



HAL
open science

Covert Capacity of Non-Coherent Rayleigh-Fading Channels

Mehrdad Tahmasbi, Anne Savard, Matthieu Bloch

► **To cite this version:**

Mehrdad Tahmasbi, Anne Savard, Matthieu Bloch. Covert Capacity of Non-Coherent Rayleigh-Fading Channels. IEEE Transactions on Information Theory, 2020, 66 (4), pp.1979-2005. 10.1109/TIT.2019.2956489 . hal-02441829

HAL Id: hal-02441829

<https://hal.science/hal-02441829v1>

Submitted on 26 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Covert Capacity of Non-Coherent Rayleigh-Fading Channels

Mehrdad Tahmasbi, Anne Savard, and Matthieu R. Bloch

Abstract

The covert capacity is characterized for a non-coherent fast Rayleigh-fading wireless channel, in which a legitimate user wishes to communicate reliably with a legitimate receiver while escaping detection from a warden. It is shown that the covert capacity is achieved with an amplitude-constrained input distribution that consists of a finite number of mass points including one at zero and numerically tractable bounds are provided. It is also conjectured that distributions with two mass points in fixed locations are optimal.

I. INTRODUCTION

In cognitive radio networks or adversarial communication settings, situations arise in which legitimate users may attempt to communicate covertly, in the sense of achieving a low probability of detection. Motivated by such applications, [1] proposed an information-theoretic model to study the throughput at which two users could reliably and covertly communicate over an Additive White Gaussian Noise (AWGN) channel in the presence of an adversary who observes the transmission through another noisy channel. The optimal covert communication throughput has been shown to satisfy a *square root law*, by which the maximum number of bits is on the order of \sqrt{n} bits over n uses of the channel. The square root law was subsequently established for some quantum channels [2] and proved to hold without requiring secret keys for binary symmetric channels under some channel conditions [3]. The exact pre-constant associated to the square root law, which plays the role of a *covert capacity*, has since been nearly completely characterized for point-to-point discrete and AWGN classical channels [4], [5], [6], as well as some classical-quantum channels [7], [8]. With the notable exception of [6], the covert capacity is typically derived when using the relative entropy as a proxy metric for covertness. Recent results [9] offer a more nuanced perspective and show that the optimal signaling scheme for covert communication over AWGN channels at finite length is metric-dependent; nevertheless, the present work still uses relative entropy to characterize covert capacity because of its convenient mathematical properties.

For Discrete Memoryless Channels (DMCs), the covert-capacity achieving input distribution takes the form of sparse signalling corresponding to those symbols that might arouse suspicion if transmitted, are used a fraction $1/\sqrt{n}$ of the time if n is the block length. Perhaps surprisingly, sparse signalling does *not* achieve the covert-capacity of AWGN channels [10], as the optimal coding scheme exploits instead Gaussian or Binary Phase-Shift Keying (BPSK) [4] signaling with an average power vanishing as $O(1/n)$. In other words, encoding information in the *phase* of modulation symbols together with a diffuse power is crucial for optimality. Gaussian signaling has therefore been used to further study covertness over Gaussian and wireless channels, as in [11], [12] to show the benefits of uninformed jammers, in [13] to analyze the role of randomized timing, in [14] to study the effect of randomized power allocation, and in [15] to analyze covert relaying strategies. We note that all aforementioned works exploit random Gaussian codebooks, which simplifies the covertness analysis by reducing the optimal attack to a radiometer. In contrast, we analyze covertness with non-random codebooks using the conceptual approach laid out in [5].

While Gaussian codebooks provides valuable insight into the properties of coding schemes for covert communications over AWGN channels, operating in the vanishing-power regime as suggested by the results might prove challenging. In particular, not only may phase-lock loops fail to properly track the phase of the transmitted signals but symbols with low amplitude may also be severely affected by phase noise, resulting in a significant degradation of the transmission reliability. These effects are also likely to be amplified by the presence of fading in wireless links. The objective of the present paper is to develop insight into this problem by characterizing the covert capacity of non-coherent fast Rayleigh-fading channels (Theorem III.1 in Section III), in which the phase is uniformly distributed over $[0; 2\pi[$; although no channel state information is available to the transmitter and receivers, some symbol-level synchronization is assumed.

Our analysis of the covert capacity for non-coherent channels builds upon the ideas initially developed in [16], [17] for amplitude constrained channels and extended to [18] for memoryless non-coherent Rayleigh fading channels under an average power constraint. In particular we show that an optimal covert capacity achieving input distribution is discrete, with one mass point located at zero and subject to an amplitude constraint. While the discrete nature of the distribution may not be a surprise, the fact that the location of the mass points is bounded results from the specific nature of the covertness constraint. We also conjecture that two mass points in *fixed* locations is actually optimal, which is supported by numerical results although we do not have a formal proof. Overall, our results suggest that, in the presence of phase uncertainty, sparse signaling might be an efficient modulation scheme for covert communication.

Our proof technique follows for the most part the high-level approach outlined in [16], [17], [18]; however, the covert communication constraint makes the analysis more intricate as the optimal capacity-achieving input distribution turns out depend on the block length. In particular, the converse arguments for single-letterization lead to a parameter-dependent constrained optimization problem, in which the parameter should be taken to zero as the blocklength goes to infinity (see the statement of Theorem III.1 and (82) in Section IV-B). This requires us to analyze the fine dependence of the objective function and the Lagrange multipliers as a function of a parameter using ideas from sensitivity analysis [19].

The rest of the paper is organized as follows. In Section III, we introduce the precise model for covert communication over non-coherent Rayleigh-fading channels and discuss our characterization of the covert capacity. In Section IV, we develop the proof of our main result, with the achievability proof in Section IV-A and the converse proof in Section IV-B.

II. NOTATION AND CONVENTIONS

Let $(\mathcal{S}, \mathcal{F})$ be a measurable space. When \mathcal{S} is a subset of \mathbb{R} , we always consider the σ -algebra induced by Borel sets, which converts \mathcal{S} to a measurable space. Let $f : \mathcal{S} \rightarrow \mathbb{R}$ be measurable and μ be a measure over $\mathcal{S} \subset \mathbb{R}$. We call f integrable if $\int_{\mathcal{S}} |f| d\mu < \infty$. We then denote the Lebesgue's integral by $\int_{\mathcal{S}} f(x) d\mu$. If $\mathcal{S} =]a, b[$ and μ is the Lebesgue's measure over \mathcal{S} , we denote $\int_{\mathcal{S}} f(x) d\mu = \int_a^b f(x) dx = \int_a^b f$. If μ is a probability measure, $X : \mathcal{S} \rightarrow \mathbb{R}$ is a random variable, and A is an event, we use $\mathbb{P}_{\mu}(A)$ and $\mathbb{E}_{\mu}(X)$ to denote $\mu(A)$ and $\int_{\mathcal{S}} X(s) d\mu$, respectively. When the probability measure μ is discrete, it can be characterized with a Probability Mass Function (PMF) $P : \mathcal{S} \rightarrow [0, 1]$ satisfying $\mu(A) = \sum_{s \in A} P(s)$. When the probability measure μ is continuous, it can be characterized with a Probability Density Function (PDF) $f : \mathcal{S} \rightarrow [0, \infty[$ satisfying $\mu(A) = \int_A f(s) ds$. We do not distinguish between a probability measure and its PMF or PDF (if they exist). The product of two measures μ and μ' is defined in the standard way and is denoted by $\mu \otimes \mu'$. We define the relative entropy between two probability measures μ and μ' as $\mathbb{D}(\mu \| \mu') \triangleq \mathbb{E}_{\mu} \left(\log \frac{d\mu}{d\mu'} \right)$, where $\frac{d\mu}{d\mu'}$ is the Radon-Nikodym derivative. We also define the χ_2 divergence as $\chi_2(\mu \| \mu') \triangleq \mathbb{E}_{\mu'} \left(\left(\frac{d\mu}{d\mu'} \right)^2 \right) - 1$. We define $\mathbb{I}(X; Y) \triangleq \mathbb{D}(\mu_{XY} \| \mu_X \otimes \mu_Y)$ where μ_{XY} , μ_X and μ_Y denote the probability measures associated to (X, Y) , X , and Y , respectively.

Let \mathcal{X} and \mathcal{Y} be two subsets of \mathbb{R} . A channel $w_{Y|X}$ from \mathcal{X} to \mathcal{Y} is a mapping $x \mapsto \mu_x$ where μ_x is a probability measure on \mathcal{Y} . If μ_x is always continuous, we write $w_{Y|X}(y|x)$ to denote the PDF of μ_x . If μ is a probability measure on \mathcal{X} and $w_{Y|X} : x \mapsto \mu'_x$ is a channel from \mathcal{X} to \mathcal{Y} , we define a joint probability measure $w_{Y|X} \times \mu$ on $\mathcal{X} \times \mathcal{Y}$ as

$$(\mu \times w_{Y|X})(\mathcal{E}) \triangleq \int \mu'_x(\mathcal{E}_x) d\mu, \quad (1)$$

where $\mathcal{E}_x \triangleq \{(\tilde{x}, \tilde{y}) \in \mathcal{E} : \tilde{x} = x\}$. We also define the marginal probability measure induced on \mathcal{Y} by $w_{Y|X} \circ \mu$. If X and Y denote the joint random variables associated to the measure $\mu \times w_{Y|X}$, we allow ourselves to denote their mutual information by $I(\mu, w_{Y|X}) \triangleq \mathbb{I}(X; Y)$.

We shall use the standard asymptotic notations such as $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, $\omega(\cdot)$ and $\Theta(\cdot)$.

III. SYSTEM MODEL AND NOTATIONS

We consider the fast Rayleigh-fading wireless channel illustrated in Fig. 1, in which at every time instant, the input-output relationships are given by

$$Y = H_m X + N_m \quad \text{and} \quad Z = H_w X + N_w, \quad (2)$$

where X is the channel input, Y is the received signal at the legitimate receiver, and Z is the received signal at the warden attempting to detect the transmission. The fading coefficients H_m and H_w are independent complex circular Gaussian random variables with zero-mean and variances θ_m^2 and θ_w^2 , respectively. The noises N_m and N_w are also independent zero-mean complex circular random variables with variance σ_m^2 and σ_w^2 , respectively. Furthermore, we assume that the channels are stationary and memoryless. The fading coefficients are unknown to all parties, who only have access to their statistical distributions. Since the phase of the fading parameters is uniform, information can only be encoded into the magnitude of X ; additionally, $|Y|^2$ and $|Z|^2$ become sufficient statistics for detection. Hence, as shown in [18], upon re-labeling $|X|^2$ by X and the outputs $|Y|^2$ and $|Z|^2$ by Y and Z , the non-coherent channel is effectively a new memoryless channel with input and output symbols in $[0, \infty[$ and transition probabilities

$$w_{Y|X}(y|x) = \frac{1}{\theta_m^2 x + \sigma_m^2} \exp\left(-\frac{y}{\theta_m^2 x + \sigma_m^2}\right) \quad \text{and} \quad w_{Z|X}(z|x) = \frac{1}{\theta_w^2 x + \sigma_w^2} \exp\left(-\frac{z}{\theta_w^2 x + \sigma_w^2}\right). \quad (3)$$

By properly normalizing Y and Z , we can assume that $\sigma_w = \sigma_m = 1$, and by normalizing X , we can further assume that $\theta_w = 1$. Thus, we can parameterize the channel by a single parameter¹ θ_m , for which the transition probabilities are

$$p_x(y) \triangleq w_{Y|X}(y|x) = \frac{1}{\theta_m^2 x + 1} \exp\left(-\frac{y}{\theta_m^2 x + 1}\right) \quad \text{and} \quad q_x(z) \triangleq w_{Z|X}(z|x) = \frac{1}{x + 1} \exp\left(-\frac{z}{x + 1}\right). \quad (4)$$

¹Note that θ_m in (3) is different from θ_m in (4).

Although the input and output sets of the channels are all equal to $[0, \infty[$, we distinguish them with the labels \mathcal{X} , \mathcal{Y} , and \mathcal{Z} for the input set, the output of main channel, and the output of the warden's channel, respectively.

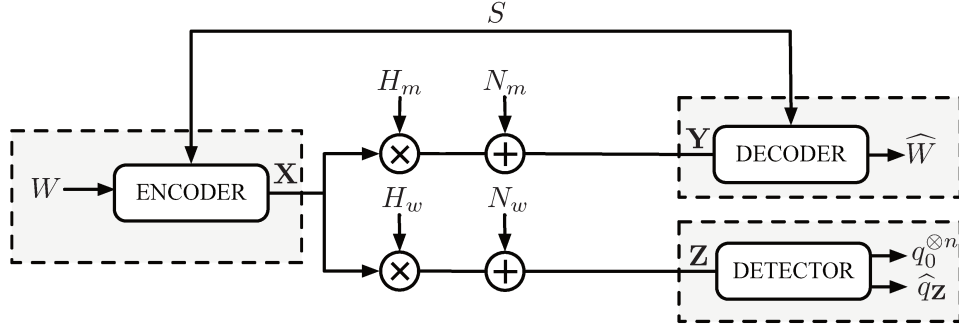


Fig. 1. Covert Wireless Channel

We next formally describe the covert communication problem in the wireless setting; as depicted in Fig. 1, the transmitter aims to communicate a message $W \in \llbracket 1, M_n \rrbracket$ by encoding it into a sequence $\mathbf{X} = (X_1, \dots, X_n)$ of n symbols using a publicly known coding scheme. Upon observing the corresponding noisy sequence $\mathbf{Y} = (Y_1, \dots, Y_n)$, the receiver forms an estimate \widehat{W} of W . The encoding and decoding may also use a pre-shared secret key S with an arbitrary distribution over a measurable space.² The objective of the warden is to detect the presence of a transmission based on its noisy observation $\mathbf{Z} = (Z_1, \dots, Z_n)$. The requirements for reliable and covert communication may be formalized as follows. We let $\widehat{q}_{\mathbf{Z}}$ denote the output distribution induced by the coding scheme and $q_0^{\otimes n}$ the product output distribution expected in the absence of communication when the channel input is set to $x = 0$. The performance of an (M_n, n) code transmitting one of M_n message over n channel uses is then measured in terms of the average probability of error $\mathbb{P}(\widehat{W} \neq W)$ and in terms of the relative entropy $\mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n})$.^{3,4} Let $\delta > 0$. We say that a covert throughput R is δ -achievable if there exist (M_n, n) codes of increasing block length n such that

$$\log M_n = \omega(\log n), \quad \lim_{n \rightarrow \infty} \mathbb{P}(\widehat{W} \neq W) = 0, \quad \limsup_{n \rightarrow \infty} \mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n}) \leq \delta, \quad \liminf_{n \rightarrow \infty} \frac{\log M_n}{\sqrt{n \mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n})}} \geq R. \quad (5)$$

The covert capacity, $C_{\text{no-CSI}}(\delta)$, is defined as the supremum of all δ -achievable covert throughputs. Note that we do not specify δ in our terminology of achievable throughput, since it turns out that the normalization of $\log M_n$ in (5) removes the dependence on δ .

Theorem III.1. *Let $\widetilde{\Omega}^{>0}$ be the set of discrete probability measures over $]0, 1[$ with a finite number of mass points. $C_{\text{no-CSI}}(\delta)$ is independently of δ equal to*

$$\sup_{\mu \in \widetilde{\Omega}^{>0}} \sqrt{2} \frac{\mathbb{E}_{\mu}(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}}. \quad (6)$$

In addition, the following simple bounds hold:

$$\max_{\tilde{x} \in]0, 1]} \tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2)} (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})) \leq C_{\text{no-CSI}}(\delta) \leq \sqrt{2} \theta_m^2. \quad (7)$$

Theorem III.1 provides useful insight into the problem of covert communication over non-coherent channels in several regards. First, a straightforward calculation shows that $\mathbb{D}(p_x \| p_0) = \theta_m^2 x - \log(1 + \theta_m^2 x)$ and $\chi_2(w_Z | X \circ \mu \| q_0) = \mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)$. The expression in (6) is therefore a counterpart of [5, Corollary 3] and [4, Eq. (28)]. Second, Theorem III.1 shows that we may restrict the signaling schemes for covert communications to finite and amplitude bounded constellations; while the finite nature of the constellation was somewhat expected from the non-coherent nature of the channel, the bound on the amplitude of the points is perhaps more surprising as it was not imposed a priori. We numerically evaluate and plot in Fig. 2 (6) when

²We show in our achievability proof that a key uniformly distributed over a discrete set with size $O(M_n)$ is sufficient to achieve the covert capacity.

³The constraint $\mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n}) \leq \delta$ ensures that, regardless of the test performed by the adversary, the sum of the probability of missed detection and false alarm is lower-bounded by $1 - \sqrt{\delta}$. Please refer to [5, Appendix A] for a detailed discussion of the operational meaning of an upper-bound on the relative entropy.

⁴The choice of this specific relative entropy to measure covertness is driven in part by the ease of analysis using channel resolvability techniques. One could of course consider alternative metrics, such as variational distance or a relative entropy with a reversed order of arguments, as discussed in [5], [6]. While the operational meaning of these other metrics remains the same, the analysis and the exact dependence on the constraint δ is metric-specific.

the number of mass points in μ is fixed using a brute-force search. Based on our numerical results, we conjecture that two mass points and On-Off Keying (OOK) signaling is optimal for covert communication.

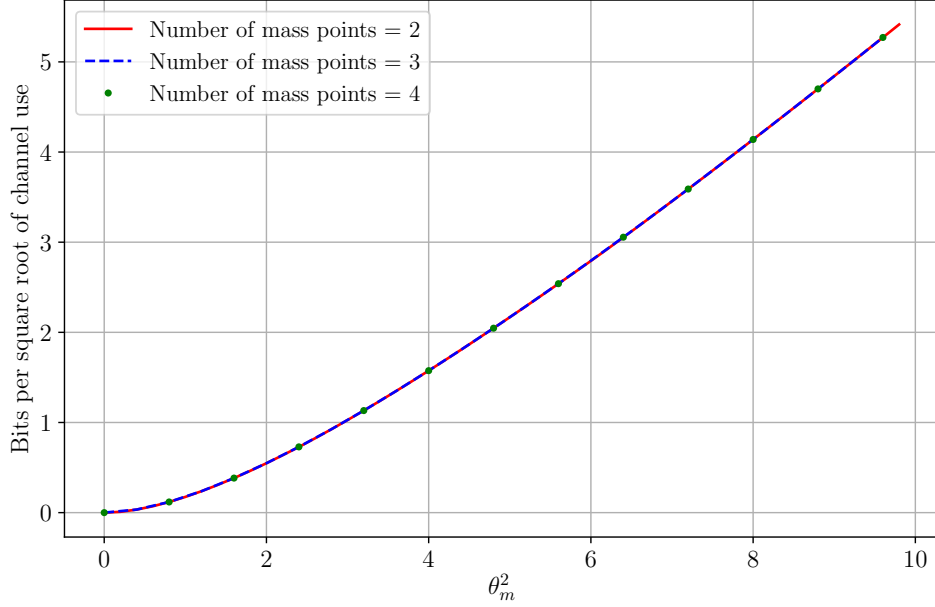


Fig. 2. Numerical evaluation of bounds on covert capacity.

IV. PROOF OF THEOREM III.1

A. Achievability proof

We prove the achievability result in two steps.

- 1) Let $\{\mu_n\}_{n \geq 1}$ be a sequence of probability measures over \mathcal{X} such that for all n , (i) for some $\tilde{x} > 0$, $\text{sup}(\text{support}(\mu_n)) \leq \tilde{x}$; and, (ii) $\limsup_{n \rightarrow \infty} n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) = \delta$. (iii) $nI(\mu_n, w_{Y|X}) = \omega(\log n)$. We then show for all $\zeta > 0$ that the cover throughput

$$(1 - \zeta) \liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}} \quad (8)$$

is δ -achievable.

- 2) Let $\mu \in \Omega^{>0}$. We construct for an arbitrary $\delta > 0$, a sequence $\{\mu_n\}_{n \geq 1}$ satisfying

$$\liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}} = \sqrt{2} \frac{\mathbb{E}_\mu(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}}, \quad (9)$$

in addition to the conditions of step 1.

1) *Step one: a random coding argument:* Although we pursue the same approach as in [5], [20] in this step, the result requires a proof of its own because of the continuous nature of the channels. Let $\{\mu_n\}_{n \geq 1}$ be a sequence of probability measures as described earlier, i.e., for all n , (i) for some $\tilde{x} > 0$, $\text{sup}(\text{support}(\mu_n)) \leq \tilde{x}$; and, (ii) $\limsup_{n \rightarrow \infty} n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) = \delta$. (iii) $I(\mu_n, w_{Y|X})n = \omega(\log n)$ For any $\zeta > 0$, we shall prove the existence of a sequence of codes $\{\mathcal{C}_n\}_{n \geq 1}$ achieving the covert throughput $(1 - \zeta) \liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}}$ with the relative entropy constraint δ . We use a random coding argument

and in particular, fix some n , and consider a random encoder $F : [1, K_n] \times [1, M_n] \rightarrow \mathcal{X}^n$ whose codewords are independent and identically distributed (i.i.d.) according to $\mu_n^{\otimes n}$. The transmitter uses the message W and the shared key S together with the encoder F to obtain the codeword $F(S, W)$ that is transmitted through the channel. By [21], for any $\gamma > 0$, we upper-bound the expected value with respect to random coding of the probability of error of an optimal decoder by

$$\mathbb{E}_F \left(\mathbb{P} \left(W \neq \widehat{W} \right) \right) \leq \mathbb{P}_{w_{Y|X}^{\otimes n} \times \mu_n^{\otimes n}} \left(\log \frac{w_{Y|X}^{\otimes n}(\mathbf{Y}|\mathbf{X})}{(w_{Y|X}^{\otimes n} \circ \mu_n^{\otimes n})(\mathbf{Y})} \geq \gamma \right) + M_n e^{-\gamma}. \quad (10)$$

Applying a Chernoff bound to the first term of the right hand side of the above inequality, for all $s > 0$, we obtain

$$\mathbb{P}_{w_{Y|X}^{\otimes n} \times \mu_n^{\otimes n}} \left(\log \frac{w_{Y|X}^{\otimes n}(\mathbf{Y}|\mathbf{X})}{(w_{Y|X}^{\otimes n} \circ \mu_n^{\otimes n})(\mathbf{Y})} \geq \gamma \right) \leq \left(\mathbb{E}_{w_{Y|X} \times \mu_n} \left(\left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu_n)(Y)} \right)^s \right) \right)^n \exp(-s\gamma). \quad (11)$$

For any probability measure μ on \mathcal{X} , upon defining

$$\phi_{\text{rel}}(s, \mu) \triangleq -\log \left(\mathbb{E}_{w_{Y|X} \times \mu} \left(\left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right)^s \right) \right), \quad (12)$$

we can re-write the right-hand side of (11) as

$$\exp(-n\phi_{\text{rel}}(s, \mu_n) - s\gamma). \quad (13)$$

To upper-bound the above expression, we need the following technical lemma describing the behavior of $\phi_{\text{rel}}(s, \mu)$ for small s .

Lemma IV.1. *For all $\tilde{x} > 0$, there exist constants $B > 0$, $\tilde{s} > 0$, and $\tilde{A} > 0$, such that for all probability measures μ and $\nu > 0$ with $\text{sup}(\text{support}(\mu)) \leq \tilde{x}$ and $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$, all $s \in [0, \tilde{s}]$, and all $A \in [\tilde{A}, \infty)$, we have*

$$\phi_{\text{rel}}(s, \mu) \geq -sI(\mu, w_{Y|X}) - B \left((2\sqrt{\nu} + \nu)e^{2A\tilde{x}+2A}A^2 + A^2e^{-A} \right) s^2 + s^3. \quad (14)$$

Proof. See Appendix D. \square

Applying Lemma IV.1 to (13), we upper-bound (11) by

$$\exp \left(-n \left(-sI(\mu_n, w_{Y|X}) - B \left(\left(2\sqrt{\frac{\delta}{n}} + \frac{\delta}{n} \right) e^{2A\tilde{x}+2A}A^2 + A^2e^{-A} \right) s^2 + s^3 \right) \right) - s\gamma. \quad (15)$$

For n large enough, we then set $A = \log n / (2(4\tilde{x} + 1))$ to ensure

$$B \left(\left(2\sqrt{\frac{\delta}{n}} + \frac{\delta}{n} \right) e^{2A\tilde{x}+2A}A^2 + A^2e^{-A} \right) = O \left(A^2 \left(n^{-\frac{1}{2}} e^{2A\tilde{x}+2A} e^{-A} \right) \right) \quad (16)$$

$$= O \left(\log^2 n \left(n^{-\frac{1}{2} + \frac{2\tilde{x}+2}{2(4\tilde{x}+1)}} + n^{-\frac{1}{2(4\tilde{x}+1)}} \right) \right) \quad (17)$$

$$= O \left(n^{-\frac{1}{2(4\tilde{x}+1)}} \log^2 n \right), \quad (18)$$

where the constant hidden in $O(\cdot)$ depends on \tilde{x} , δ , and the channel. Therefore, we have for $s = n^{-\beta}$,

$$B \left(\left(\left(2\sqrt{\frac{\delta}{n}} + \frac{\delta}{n} \right) e^{2A\tilde{x}+2A}A^2 + A^2e^{-A} \right) s^2 + s^3 \right) \stackrel{(a)}{=} O \left(n^{-\frac{1}{2(4\tilde{x}+1)}} \log^2 n s^2 + s^3 \right) \quad (19)$$

$$= O \left(n^{-\frac{1}{2(4\tilde{x}+1)} - 2\beta} \log^2 n + n^{-3\beta} \right) \quad (20)$$

$$= O \left(n^{-\min(3\beta, 2\beta+1/(2(4\tilde{x}+1)))} \log^2 n \right), \quad (21)$$

where (a) follows from (18). The expression in (21) will be $o(sI(\mu_n, w_{Y|X}))$ when $I(\mu_n, w_{Y|X}) = \Omega \left(n^{-\frac{1}{2}} \right)$ and $\max(1/4, 1/2 - 1/(2(4\tilde{x} + 1))) < \beta$. Moreover, if we choose $\beta < 1/2$, which is feasible with the previous constraint, we guarantee that $snI(\mu_n, w_{Y|X}) \geq n^c$ for some $c > 0$ and n large enough. Finally, for $\gamma = (1 - \zeta/2)I(\mu_n, w_{Y|X})n$ and $\log M_n = (1 - \zeta)I(\mu_n, w_{Y|X})n$, we have by (10)

$$\mathbb{E}_F \left(\mathbb{P} \left(W \neq \widehat{W} \right) \right) \leq \exp \left(-(1 + o(1)) \frac{\zeta}{2} sI(\mu_n, w_{Y|X})n \right) + \exp \left(-\frac{\zeta}{2} I(\mu_n, w_{Y|X})n \right) \quad (22)$$

$$\leq 2 \exp(-\zeta n^c). \quad (23)$$

This completes the reliability part of the proof.

We now proceed to the resolvability part. Recall that we denote the induced distribution at the output of the warden's channel by $\widehat{p}_{\mathbf{Z}} \triangleq \frac{1}{M_n K_n} \sum_{s=1}^{K_n} \sum_{w=1}^{M_n} w_{Z|X}^{\otimes n}(\mathbf{z}|F(s, w))$, where M_n and K_n are the message size and the key size, respectively. By a modification of [22, Equation (194)], we know that for all $s \in [0, 1]$,

$$\mathbb{E}_F \left(\mathbb{D} \left(\widehat{p}_{\mathbf{Z}} \| (w_{Z|X} \circ \mu_n)^{\otimes n} \right) \right) \leq \frac{1}{s} \exp(-s \log(M_n K_n) - n\phi_{\text{res}}(s, \mu_n)), \quad (24)$$

where

$$\phi_{\text{res}}(s, \mu) \triangleq -\log \left(\mathbb{E}_{w_{Z|X} \times \mu} \left(\left(\frac{q_X(Z)}{(w_{Z|X} \circ \mu)(Z)} \right)^s \right) \right). \quad (25)$$

Since the above function is the same as ϕ_{rel} except that $w_{Y|X}$ is replaced by $w_{Z|X}$, $w_{Z|X}$ is a special case of $w_{Y|X}$ for $\theta_m = 1$, and we choose s in the reliability part so that $\log \frac{1}{s} = O(\log n)$, we can follow the same approach to obtain for some $\tilde{c} > 0$,

$$\mathbb{E}_F \left(\mathbb{D} \left(\hat{p}_{\mathbf{Z}} \| (w_{Z|X} \circ \mu_n)^{\otimes n} \right) \right) \leq 2 \exp \left(-n^{\tilde{c}} \right), \quad (26)$$

if $\log M_n + \log K_n \geq (1 + \zeta)I(\mu_n, w_{Z|X})n$. Therefore, the expected value of the covertness of the random code is

$$\mathbb{E}_F \left(\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \right) = \mathbb{E}_F \left(\int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \log \frac{\hat{p}_{\mathbf{Z}}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \right) \quad (27)$$

$$= \mathbb{E}_F \left(\int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \log \frac{\hat{p}_{\mathbf{Z}}(\mathbf{z})}{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})} d\mathbf{z} + \int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \right) \quad (28)$$

$$= \mathbb{E}_F \left(\mathbb{D}(\hat{p}_{\mathbf{Z}} \| (w_{Z|X} \circ \mu_n)^{\otimes n}) \right) + \mathbb{E}_F \left(\int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \right) \quad (29)$$

$$\leq 2 \exp \left(-n^{\tilde{c}} \right) + \mathbb{E}_F \left(\int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \right) \quad (30)$$

$$\stackrel{(a)}{=} 2 \exp \left(-n^{\tilde{c}} \right) + \int_{\mathbb{R}^n} \mathbb{E}_F(\hat{p}_{\mathbf{Z}}(\mathbf{z})) \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \quad (31)$$

$$= 2 \exp \left(-n^{\tilde{c}} \right) + \int_{\mathbb{R}^n} (w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z}) \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} d\mathbf{z} \quad (32)$$

$$= 2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0), \quad (33)$$

where (a) follows from Fubini's theorem and $\mathbb{E}_F \left(\int_{\mathbb{R}^n} \hat{p}_{\mathbf{Z}}(\mathbf{z}) \left| \log \frac{(w_{Z|X} \circ \mu_n)^{\otimes n}(\mathbf{z})}{q_0^{\otimes n}(\mathbf{z})} \right| d\mathbf{z} \right) < \infty$ by Lemma C.4. Applying Markov's inequality, for large n , we obtain

$$\mathbb{P}_F \left(\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \leq \frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right), \mathbb{P} \left(W \neq \widehat{W} \right) \leq 4n \exp \left(-n^c \right) \right) \quad (34)$$

$$\geq 1 - \mathbb{P}_F \left(\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \geq \frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right) \right) - \mathbb{P}_F \left(\mathbb{P} \left(W \neq \widehat{W} \right) \geq 4n \exp \left(-n^c \right) \right) \quad (35)$$

$$\geq 1 - \frac{\mathbb{E}_F(\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}))}{\frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right)} - \frac{\mathbb{E}_F \left(\mathbb{P} \left(W \neq \widehat{W} \right) \right)}{4n \exp \left(-n^c \right)} \quad (36)$$

$$\geq 1 - \frac{2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}{\frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right)} - \frac{2 \exp \left(-n^c \right)}{4n \exp \left(-n^c \right)} \quad (37)$$

$$= 1 - \frac{n}{n+1} - \frac{1}{2n} > 0. \quad (38)$$

This implies that there exists a sequence of codes $\{\mathcal{C}_n\}_{n \geq 1}$ such that \mathcal{C}_n satisfies

$$\log M_n = (1 - \zeta)I(\mu_n, w_{Y|X})n = \omega(\log n), \quad (39)$$

$$\log M_n + \log K_n = (1 + \zeta)I(\mu_n, w_{Z|X})n, \quad (40)$$

$$P_e \leq 4n \exp \left(-n^c \right), \quad (41)$$

$$\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \leq \frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right). \quad (42)$$

The covert throughput would be then

$$\liminf_{n \rightarrow \infty} \frac{\log M_n}{\sqrt{n\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n})}} = \liminf_{n \rightarrow \infty} \frac{(1 - \zeta)I(\mu_n, w_{Y|X})n}{\sqrt{n\mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n})}} \quad (43)$$

$$\geq \liminf_{n \rightarrow \infty} \frac{(1 - \zeta)I(\mu_n, w_{Y|X})n}{\sqrt{n \left(\frac{n+1}{n} \left(2 \exp \left(-n^{\tilde{c}} \right) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right) \right)}} \quad (44)$$

$$= \liminf_{n \rightarrow \infty} \frac{(1 - \zeta)I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}}. \quad (45)$$

Since $\limsup_{n \rightarrow \infty} n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \delta$ by our assumption, we have

$$\limsup_{n \rightarrow \infty} \mathbb{D}(\widehat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \leq \limsup_{n \rightarrow \infty} \left(\frac{n+1}{n} \left(2 \exp(-n\tilde{c}) + n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \right) \right) \leq \delta. \quad (46)$$

2) *Step two: obtaining the bound in Theorem III.1:* Let $\mu \in \tilde{\Omega}^{>0}$ and μ_0 be the probability measure with a single mass point at zero. We define $\alpha_n \triangleq \sqrt{\frac{\delta}{n\chi_2(w_{Z|X} \circ \mu \| q_0)}}$ and $\mu_n \triangleq \alpha_n \mu + (1-\alpha_n)\mu_0$. We have $\max(\text{support}(\mu_n)) = \max(\text{support}(\mu)) \triangleq a < 1$ by definition of $\tilde{\Omega}^{>0}$. Hence, it is enough to check that

$$nI(\mu_n, w_{Y|X}) = \omega(\log n), \quad (47)$$

$$\limsup_{n \rightarrow \infty} n\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \delta, \quad (48)$$

$$\liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}} \geq \sqrt{2} \frac{\mathbb{E}_\mu(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}}. \quad (49)$$

We next state a lemma providing a general upper-bound for the relative entropy in terms of the χ_2 divergence.

Lemma IV.2. *Let $\mu \in \tilde{\Omega}^{\geq 0}$ with $\max(\text{support}(\mu)) \leq a < 1$. Let $M > 0$ and $\epsilon > 0$. We have*

$$\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \frac{1}{2} \chi_2(w_{Z|X} \circ \mu \| q_0) + (\mathbb{E}_\mu(X))^3 + (\mathbb{E}_\mu(X))^4 \int_0^M e^{z(-1 + \frac{4a}{1+a})} dz + \int_M^\infty e^{-\frac{z}{1+\epsilon}} z dz + \frac{\mathbb{E}_\mu(X)}{\epsilon} \int_M^\infty e^{-\frac{z}{1+a}} z dz \quad (50)$$

Proof. See Appendix E. □

Applying Lemma IV.2 to μ_n with some M_n and ϵ , we obtain

$$\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \frac{1}{2} \chi_2(w_{Z|X} \circ \mu_n \| q_0) + (\mathbb{E}_{\mu_n}(X))^3 + (\mathbb{E}_{\mu_n}(X))^4 \int_0^{M_n} e^{z(-1 + \frac{4a}{1+a})} dz + \int_{M_n}^\infty e^{-\frac{z}{1+\epsilon}} z dz + \frac{1}{\epsilon} \mathbb{E}_{\mu_n}(X) \int_{M_n}^\infty e^{-\frac{z}{1+a}} z dz, \quad (51)$$

where $a = \max(\text{support}(\mu))$. We will prove for appropriately chosen M_n and ϵ that

$$(\mathbb{E}_{\mu_n}(X))^3 + (\mathbb{E}_{\mu_n}(X))^4 \int_0^{M_n} e^{z(-1 + \frac{4a}{1+a})} dz + \int_{M_n}^\infty e^{-\frac{z}{1+\epsilon}} z dz + \frac{1}{\epsilon} \mathbb{E}_{\mu_n}(X) \int_{M_n}^\infty e^{-\frac{z}{1+a}} z dz = o(\alpha_n^2) \quad (52)$$

Note that $\mathbb{E}_{\mu_n}(X) = \alpha_n \mathbb{E}_\mu(X)$, and therefore, $(\mathbb{E}_{\mu_n}(X))^3 = O(\alpha_n^3) = o(\alpha_n^2)$. We choose $M_n = B \log \frac{1}{\alpha_n}$, where B is a constant independent of n specified later. We then have

$$(\mathbb{E}_{\mu_n}(X))^4 \int_0^{M_n} e^{z(-1 + \frac{4a}{1+a})} dz \leq \begin{cases} O\left(\alpha_n^4 M_n e^{M_n(-1 + \frac{4a}{1+a})}\right) & a > 1/3 \\ O\left(\alpha_n^4 M_n\right) & a \leq 1/3 \end{cases} \quad (53)$$

$$= \begin{cases} O\left(\alpha_n^{4-B\frac{3a-1}{a+1}} \log \frac{1}{\alpha_n}\right) & a > 1/3 \\ O\left(\alpha_n^4 \log \frac{1}{\alpha_n}\right) & a \leq 1/3 \end{cases} \quad (54)$$

$$\stackrel{(a)}{=} o(\alpha_n^2), \quad (55)$$

where (a) requires that $B < 2\frac{1+a}{3a-1}$ when $a > 1/3$. We further have

$$\int_{M_n}^\infty e^{-\frac{z}{1+\epsilon}} z dz = (1+\epsilon)^2 e^{-\frac{M_n}{1+\epsilon}} \left(\frac{M_n}{1+\epsilon} + 1 \right) \quad (56)$$

$$= (1+\epsilon)^2 \alpha_n^{\frac{B}{1+\epsilon}} \left(\frac{B \log \frac{1}{\alpha_n}}{1+\epsilon} + 1 \right) \quad (57)$$

$$\stackrel{(a)}{=} o(\alpha_n^2), \quad (58)$$

where (a) requires that $B > 2(1 + \epsilon)$. Finally, we have

$$\frac{1}{\epsilon} \mathbb{E}_{\mu_n}(X) \int_{M_n}^{\infty} e^{-\frac{z}{1+a}} z dz = \frac{1}{\epsilon} \mathbb{E}_{\mu_n}(X) (1+a)^2 e^{-\frac{M_n}{1+a}} \left(\frac{M_n}{1+a} + 1 \right) \quad (59)$$

$$= \frac{1}{\epsilon} \mathbb{E}_{\mu_n}(X) (1+a)^2 \alpha_n^{\frac{B}{1+a}} \left(\frac{B \log \frac{1}{\alpha_n}}{1+a} + 1 \right) \quad (60)$$

$$\stackrel{(a)}{=} o(\alpha_n^2), \quad (61)$$

where (a) requires that $B > 1 + a$. If $a \leq 1/3$, we only need to choose B and ϵ such that $B > \max(2(1 + \epsilon), 1 + a)$. For $a > 1/3$, we choose $0 < \epsilon < \frac{1+a}{3a-1} - 1$ so that $\max(2(1 + \epsilon), 1 + a) < 2 \frac{1+a}{3a-1}$. We then choose B such that $\max(2(1 + \epsilon), 1 + a) < B < 2 \frac{1+a}{3a-1}$. This complete the proof of (52). Note next that by Lemma C.5

$$\chi_2(w_{Z|X} \circ \mu_n \| q_0) = \mathbb{E}_{\mu_n \otimes \mu_n} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right) \quad (62)$$

$$= \mathbb{E}_{\mu_n \otimes \mu_n} \left(\frac{X_1 X_2}{1 - X_1 X_2} | X_1 > 0, X_2 > 0 \right) \mathbb{P}_{\mu_n \otimes \mu_n}(X_1 > 0, X_2 > 0) \quad (63)$$

$$\stackrel{(a)}{=} \alpha_n^2 \mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right) \quad (64)$$

$$= \alpha_n^2 \chi_2(w_{Z|X} \circ \mu \| q_0) \quad (65)$$

$$\stackrel{(b)}{=} \frac{\delta}{n}, \quad (66)$$

where (a) follows from the definition of μ_n and (b) follows from the definition of α_n . We therefore have

$$\limsup_{n \rightarrow \infty} n \mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \delta. \quad (67)$$

Following the same reasoning, one can show that $\mathbb{D}(w_{Y|X} \circ \mu_n \| p_0) = O(\alpha_n^2)$. Finally, we have

$$I(\mu_n, w_{Y|X}) = \mathbb{E}_{\mu_n}(\theta_m^2 X - \log(1 + \theta_m^2 X)) - \mathbb{D}(w_{Y|X} \circ \mu_n \| p_0) \quad (68)$$

$$= \mathbb{E}_{\mu_n}(\theta_m^2 X - \log(1 + \theta_m^2 X)) - O(\alpha_n^2) \quad (69)$$

$$= \alpha_n \mathbb{E}_{\mu}(\theta_m^2 X - \log(1 + \theta_m^2 X)) - O(\alpha_n^2) \quad (70)$$

$$= \Omega(n^{-\frac{1}{2}}) \quad (71)$$

$$= \omega\left(\frac{\log n}{n}\right), \quad (72)$$

which yields that

$$\liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}} \geq \sqrt{2} \frac{\mathbb{E}_{\mu}(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}}. \quad (73)$$

To obtain the lower-bound in (7), we choose μ to be a probability measure with a single mass point at $\tilde{x} \in]0, 1[$. We then have

$$\sqrt{2} \frac{\mathbb{E}_{\mu}(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}} = \tilde{x}^{-1} \sqrt{2(1 - \tilde{x})} (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})). \quad (74)$$

B. Converse proof

Before delving into the detailed proofs, we first provide the sketch of the various steps of the converse proof.

- 1) We first follow the reasoning of the converse proof of [5] to show that if R is a δ -achievable rate, then there exists a sequence of probability measures $\{\mu_n\}_{n \geq 1}$ over \mathcal{X} such that $\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \delta/n$ for n and

$$R \leq \liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}}. \quad (75)$$

- 2) We show that the probability measure μ_n can be further restricted to be discrete with a finite number of mass points and a mass point at zero. This is achieved by investigating the optimization problem

$$\sup_{\mu: \mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu} I(\mu, w_{Y|X}), \quad (76)$$

and adapting some techniques developed in [18].

- 3) We prove that we can still upper-bound a covert throughput even if we constraint the amplitude of μ_n as $\max(\text{support}(\mu_n)) \leq 1 + \zeta$ for any $\zeta > 0$.
- 4) Let $\{\mu_n\}_{n \geq 1}$ be a sequence of probability measures such that μ_n has a finite number of mass and $\max(\text{support}(\mu_n)) \leq 1 + \zeta$. We show that

$$\liminf_{n \rightarrow \infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0)}} \leq \sup_{\mu \in \Omega^{>0}} \frac{\sqrt{2} \mathbb{E}_\mu(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right)}}. \quad (77)$$

1) *Step one: a general converse for covert communication.* We consider a sequence of code $\{\mathcal{C}_n\}_{n \geq 1}$ where each code \mathcal{C}_n can transmit $\log M_n$ bits with probability of error ϵ_n and relative entropy at most δ_n , and we have $\lim_{n \rightarrow \infty} \epsilon_n = 0$ and $\limsup_{n \rightarrow \infty} \delta_n \leq \delta$. If $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ denotes the input and the output of the channels when \mathcal{C}_n is used and $\hat{p}_{\mathbf{X}\mathbf{Y}\mathbf{Z}}$ denotes the joint distribution, a standard application of Fano's inequality yields

$$\log M_n \leq \frac{\mathbb{I}(\mathbf{X}; \mathbf{Y}) + \mathbb{H}_b(\epsilon_n)}{1 - \epsilon_n} \leq \frac{\mathbb{I}(\mathbf{X}; \mathbf{Y}) + 1}{1 - \epsilon_n}, \quad (78)$$

where $\mathbb{H}_b(x) \triangleq -x \log(x) - (1-x) \log(1-x)$. One can then upper-bound the mutual information $\mathbb{I}(\mathbf{X}; \mathbf{Y})$ using standard techniques [23] to obtain

$$\mathbb{I}(\mathbf{X}; \mathbf{Y}) \leq \sum_{i=1}^n \mathbb{I}(X_i; Y_i) \leq n \mathbb{I}(\tilde{X}_n; \tilde{Y}_n), \quad (79)$$

where the random variables \tilde{X}_n and \tilde{Y}_n are distributed according to $p_{\tilde{X}_n}(x) \triangleq \frac{1}{n} \sum_{i=1}^n \hat{p}_{X_i}(x)$ and $p_{\tilde{X}_n \tilde{Y}_n}(x, y) \triangleq p_{\tilde{X}_n}(x) p_x(y)$. Note that $\lim_{n \rightarrow \infty} n \mathbb{I}(\tilde{X}_n; \tilde{Y}_n) = \infty$ since we assumed that $\log M_n = \omega(\log n)$. Following [24], [4], one can also lower-bound the relative entropy as

$$\delta_n \geq \mathbb{D}(\hat{p}_{\mathbf{Z}} \| q_0^{\otimes n}) \geq \sum_{i=1}^n \mathbb{D}(\hat{p}_{Z_i} \| q_0) \geq n \mathbb{D}(p_{\tilde{Z}_n} \| q_0), \quad (80)$$

where \tilde{Z}_n is distributed according to $p_{\tilde{Z}_n}(z) \triangleq \frac{1}{n} \sum_{i=1}^n \hat{p}_{Z_i}(z)$. Consequently,

$$C_{\text{no-CSI}} \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{I}(\tilde{X}_n; \tilde{Y}_n)}{(1 - \epsilon_n) \sqrt{\mathbb{D}(p_{\tilde{Z}_n} \| q_0)}} \left(1 + \frac{1}{n \mathbb{I}(\tilde{X}_n; \tilde{Y}_n)} \right) = \liminf_{n \rightarrow \infty} \frac{\mathbb{I}(\tilde{X}_n; \tilde{Y}_n)}{\sqrt{\mathbb{D}(p_{\tilde{Z}_n} \| q_0)}} \quad (81)$$

where the sequence of distributions $\{p_{\tilde{X}_n \tilde{Y}_n \tilde{Z}_n}\}_{n \geq 0}$ is subject to the constraint $\mathbb{D}(p_{\tilde{Z}_n} \| q_0) \leq \frac{\delta_n}{n}$. This completes the first step of the converse proof.

2) *Step two: discreteness of the optimal distribution.* We define the optimization problem

$$A(\nu) \triangleq \sup_{\mu \in \Omega: \mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu} I(\mu, w_{Y|X}), \quad (82)$$

where Ω is the set of all probability measures over \mathcal{X} such as μ such that $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) < \infty$. The next lemma shows that there exists a unique maximizer to the above problem.

Lemma IV.3. *Let $\nu > 0$. There exists a unique probability measure $\mu_\nu^* \in \Omega$ such that $\mathbb{D}(w_{Z|X} \circ \mu_\nu^* \| q_0) \leq \nu$ and $I(\mu_\nu^*, w_{Y|X}) = A(\nu)$.*

Proof. See Appendix F. □

We next characterize the unconstrained form of the optimization in (82).

Theorem IV.1. *Let $\nu > 0$. There exists $\gamma(\nu) \geq 0$ such that the following holds.*

- 1) *We have*

$$A(\nu) = \max_{\mu \in \Omega} [I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)], \quad (83)$$

and μ_ν^* is the unique maximizer of the above optimization.

2) Define

$$w(x, \mu_1, \nu) \triangleq \int_0^\infty p_x(y) \log \frac{p_x(y)}{(w_{Y|X} \circ \mu_1)(y)} dy - \gamma(\nu) \left(\int_0^\infty q_x(z) \log \frac{(w_{Z|X} \circ \mu_1)(z)}{q_0(z)} dz - \nu \right). \quad (84)$$

For all $\mu \in \Omega$, we have

$$A(\nu) \geq \mathbb{E}_\mu(w(X, \mu_\nu^*, \nu)). \quad (85)$$

3) Given $\mu_1 \in \Omega$, we have for all $\mu \in \Omega$,

$$A(\nu) \geq \mathbb{E}_\mu(w(X, \mu_1, \nu)). \quad (86)$$

if and only if

$$w(x, \mu_1, \nu) \leq A(\nu) \quad \forall x \in \mathcal{X}, \quad (87)$$

$$w(x, \mu_1, \nu) = A(\nu) \quad \forall x \in \text{support}(\mu_1). \quad (88)$$

4) We have $\lim_{\nu \rightarrow 0^+} \gamma(\nu) = \infty$ and $\lim_{\nu \rightarrow 0^+} \gamma(\nu)\nu = 0$.

Proof. See Appendix F. □

Lemma IV.4. *There exists $\nu_0 > 0$ such that for all $0 < \nu \leq \nu_0$, $\text{support}(\mu_\nu^*)$ is discrete with a finite number of points in any bounded interval.*

Proof. Fix some $\nu > 0$, and define $r(y) \triangleq (w_{Y|X} \circ \mu_\nu^*)(y)$ and $f(z) \triangleq (w_{Z|X} \circ \mu_\nu^*)(z)$. We assume that there exists an interval with an infinite number of points in $\text{support}(\mu_\nu^*)$ and obtain a contradiction for ν small enough in four steps.

Step 1: We first use the argument in [18] to show that the KKT condition in (88) holds for all $x \geq 0$. By the Bolzano-Weierstrass theorem, there exists a convergent sequence $\{x_i\}_{i \geq 1}$ in $\text{support}(\mu_\nu^*)$. Moreover, by (88), for any $x \in \text{support}(\mu_\nu^*)$, we have

$$\phi_\nu(x) \triangleq w(x, \mu_\nu^*, \nu) - A(\nu) \quad (89)$$

$$= \int_0^\infty p_x(y) \log \frac{p_x(y)}{r(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{f(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu = 0. \quad (90)$$

We now show that $\phi_\nu(x)$ is analytic in x over the domain $\mathcal{D} \triangleq \{x : \mathcal{R}(x) > 0\}$. Note that $\int_0^\infty p_x(y) \log p_x(y) dy = -\log(1 + \theta_m x) - 1$ and $\int_0^\infty q_x(z) \log q_0(z) dz = -1 - x$, which are analytic over \mathcal{D} . We furthermore have

$$|p_x(y)| = \frac{1}{|1 + \theta_m^2 x|} \left| e^{-\frac{y}{1 + \theta_m^2 x}} \right| \quad (91)$$

$$\stackrel{(a)}{=} \frac{1}{|1 + \theta_m^2 x|} e^{-\frac{y(\theta_m^2 \mathcal{R}(x) + 1)}{|1 + \theta_m^2 x|^2}}, \quad (92)$$

where (a) follows from $|e^z| = e^{\mathcal{R}(z)}$. This implies that

$$\int_0^\infty |p_x(y) \log r(y)| dy \stackrel{(a)}{\leq} \int_0^\infty |p_x(y)| (\theta_m^2 \mathbb{E}_{\mu_\nu^*}(X) + y) dy \quad (93)$$

$$\leq \int_0^\infty |p_x(y)| (\theta_m^2 (2\sqrt{\nu} + \nu) + y) dy \quad (94)$$

$$\stackrel{(c)}{=} \theta_m^2 (2\sqrt{\nu} + \nu) \frac{|1 + \theta_m^2 x|}{\theta_m^2 \mathcal{R}(x) + 1} + \frac{|1 + \theta_m^2 x|^3}{(\theta_m^2 \mathcal{R}(x) + 1)^2}, \quad (95)$$

where (a) follows from (205), (b) follows from Lemma C.2, and (c) follows by (92). Therefore, $\int_0^\infty |p_x(y) \log r(y)| dy$ is uniformly bounded on any compact subset of \mathcal{D} , and Theorem B.1 yields that $\int_0^\infty |p_x(y) \log r(y)| dy$ is analytic over \mathcal{D} . One can similarly argue that $\int_0^\infty q_x(z) \log f(z) dz$ is also analytic over \mathcal{D} and therefore ϕ_ν is analytic. Since $\phi_\nu(x)$ is an analytic function over \mathcal{D} , and $\phi_\nu(x) = 0$ over a set with a limit point in \mathcal{D} , the identity theorem [25] states that $\phi_\nu(x) = 0$ for all $x \in \mathcal{D}$. Thus, $\phi_\nu(x) = 0$ holds over the entire real line. Using $\int_0^\infty p_x(y) \log p_x(y) dy = -\log(1 + \theta_m x) - 1$ and $\int_0^\infty q_x(z) \log q_0(z) dz = -1 - x$, we can re-write

$$0 = \phi_\nu(x) = -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1 + x) - A(\nu) + \gamma(\nu)\nu - \int_0^\infty p_x(y) \log r(y) dy - \gamma(\nu) \int_0^\infty q_x(z) \log f(z) dz. \quad (96)$$

To obtain a contradiction, we cannot use the Laplace transform approach of [18] because there are two integrals in (96), which is therefore the sum of two Laplace transforms with different arguments. Hence, we continue the proof with another approach.

Step 2: In this step, we shall find the supremum of the support of μ_ν^* in terms of $\gamma(\nu)$. We first consider any non-zero point $\tilde{x} \in \text{support}(\mu_\nu^*)$ and any $\Delta \in]0; \tilde{x}[$. Since $\tilde{x} \in \text{support}(\mu_\nu^*)$, there exists $\delta > 0$ with $\mu_\nu^*(] \tilde{x} - \Delta, \tilde{x} + \Delta]) = \delta$. Thus, for any y , by definition of $r(y)$ and the law of total probability, we lower-bound $r(y)$ by

$$r(y) = \mathbb{E}_{\mu_\nu^*} \left(\frac{1}{1 + \theta_m^2 X} e^{-\frac{y}{1 + \theta_m^2 X}} \right) \quad (97)$$

$$\geq \mathbb{E}_{\mu_\nu^*} \left(\frac{1}{1 + \theta_m^2 X} e^{-\frac{y}{1 + \theta_m^2 X}} \middle| X \in] \tilde{x} - \Delta, \tilde{x} + \Delta[\right) \mu_\nu^*(] \tilde{x} - \Delta, \tilde{x} + \Delta[) \quad (98)$$

$$\geq \frac{\delta}{1 + \theta_m^2 (\tilde{x} + \Delta)} e^{-\frac{y}{1 + \theta_m^2 (\tilde{x} + \Delta)}}, \quad (99)$$

and similarly, lower-bound $f(z)$ by

$$f(z) \geq \frac{\delta}{1 + \tilde{x} + \Delta} e^{-\frac{z}{1 + \tilde{x} + \Delta}}. \quad (100)$$

Substituting these bounds in (96), we obtain

$$\begin{aligned} 0 &\leq -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1 + x) - A(\nu) + \gamma(\nu)\nu \\ &\quad - \int_0^\infty p_x(y) \log \frac{\delta}{1 + \theta_m^2 (\tilde{x} + \Delta)} e^{-\frac{y}{1 + \theta_m^2 (\tilde{x} + \Delta)}} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{\delta}{1 + \tilde{x} + \Delta} e^{-\frac{z}{1 + \tilde{x} + \Delta}} dz \end{aligned} \quad (101)$$

$$\begin{aligned} &= -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1 + x) - A(\nu) + \gamma(\nu)\nu \\ &\quad - \log \frac{\delta}{1 + \theta_m^2 (\tilde{x} + \Delta)} + \frac{1 + \theta_m^2 x}{1 + \theta_m^2 (\tilde{x} - \Delta)} - \gamma(\nu) \left(\log \frac{\delta}{1 + \tilde{x} + \Delta} - \frac{1 + x}{1 + \tilde{x} - \Delta} \right) \end{aligned} \quad (102)$$

$$= \kappa - \log(\theta_m^2 x + 1) - x \left(\gamma(\nu) \frac{\tilde{x} - \Delta}{1 + \tilde{x} - \Delta} - \frac{\theta_m^2}{1 + \theta_m^2 (\tilde{x} - \Delta)} \right), \quad (103)$$

where κ is a constant not depending on x . Since (103) holds for all x , by taking the limit $x \rightarrow \infty$, we should have

$$\gamma(\nu) \frac{\tilde{x} - \Delta}{1 + \tilde{x} - \Delta} - \frac{\theta_m^2}{1 + \theta_m^2 (\tilde{x} - \Delta)} \leq 0. \quad (104)$$

Moreover, by letting Δ tend to zero, we obtain

$$\gamma(\nu) \frac{\tilde{x}}{1 + \tilde{x}} - \frac{\theta_m^2}{1 + \theta_m^2 \tilde{x}} \leq 0, \quad (105)$$

which implies that $x^* \triangleq \sup(\text{support}(\mu_\nu^*)) < \infty$. Furthermore, upon finiteness of x^* , we have

$$r(y) \leq e^{-\frac{y}{1 + \theta_m^2 x^*}}, \quad (106)$$

and

$$f(z) \leq e^{-\frac{z}{1 + x^*}}. \quad (107)$$

Replacing these upper-bounds in (96), we obtain

$$0 \geq -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1 + x) - A(\nu) + \gamma(\nu)\nu - \int_0^\infty p_x(y) \log e^{-\frac{y}{1 + \theta_m^2 x^*}} dy - \gamma(\nu) \int_0^\infty q_x(z) \log e^{-\frac{z}{1 + x^*}} dz \quad (108)$$

$$= -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1 + x) - A(\nu) + \gamma(\nu)\nu + \frac{1 + \theta_m^2 x}{1 + \theta_m^2 x^*} + \gamma(\nu) \frac{1 + x}{1 + x^*} \quad (109)$$

$$= \kappa' - \log(\theta_m^2 x + 1) - x \left(\gamma(\nu) \frac{x^*}{1 + x^*} - \frac{\theta_m^2}{1 + \theta_m^2 x^*} \right), \quad (110)$$

where κ' is a constant not depending on x . Since (110) holds for all x , we have

$$\gamma(\nu) \frac{x^*}{1 + x^*} - \frac{\theta_m^2}{1 + \theta_m^2 x^*} \geq 0. \quad (111)$$

By definition of the support of a distribution, it should be closed, and therefore, $x^* \in \text{support}(\mu_\nu^*)$. Since (105) holds for all points in the support, we can set $\tilde{x} = x^*$ and obtain

$$\gamma(\nu) \frac{x^*}{1 + x^*} - \frac{\theta_m^2}{1 + \theta_m^2 x^*} = 0. \quad (112)$$

Step 3: Using the equality for x^* in (112), we derive an upper-bound on $A(\nu)$ depending on $\gamma(\nu)$ and ν . By definition of μ_ν^* , it holds that

$$A(\nu) = I(\mu_\nu^*, w_{Y|X}) \quad (113)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu_\nu^*} \left(\log \frac{p_X(Y)}{r(Y)} \right) \quad (114)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu_\nu^*} \left(\log \frac{p_X(Y)p_0(Y)}{r(Y)p_0(Y)} \right) \quad (115)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu_\nu^*} \left(\log \frac{p_X(Y)}{p_0(Y)} \right) - \mathbb{E}_{w_{Y|X} \circ \mu_\nu^*} \left(\log \frac{r(Y)}{p_0(Y)} \right) \quad (116)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu_\nu^*} \left(\log \frac{p_X(Y)}{p_0(Y)} \right) - \mathbb{D}(r \| p_0) \quad (117)$$

$$\leq \mathbb{E}_{w_{Y|X} \times \mu_\nu^*} \left(\log \frac{p_X(Y)}{p_0(Y)} \right) \quad (118)$$

$$= \mathbb{E}_{\mu_\nu^*} (\theta_m^2 X - \log(1 + \theta_m^2 X)) \quad (119)$$

$$\stackrel{(a)}{\leq} \mathbb{E}_{\mu_\nu^*} \left(\frac{1}{2} \theta_m^4 X^2 \right) \quad (120)$$

$$\leq \frac{1}{2} \theta_m^4 x^* \mathbb{E}(X) \quad (121)$$

$$\stackrel{(b)}{\leq} \frac{1}{2} \theta_m^4 x^* (2\sqrt{\nu} + \nu), \quad (122)$$

where (a) follows from $\log(1+x) \geq x - x^2/2$ for $x \geq 0$, and (b) follows from Lemma C.2. Therefore, we can use (112) to obtain

$$A(\nu) \leq \frac{1}{2} \theta_m^4 \left(\frac{\theta_m^4 (1+x^*)}{\gamma(\nu)(1+\theta_m^4 x^*)} \right) (2\sqrt{\nu} + \nu) \quad (123)$$

$$\leq \frac{2\sqrt{\nu} + \nu}{\gamma(\nu)} \left(\frac{1}{2} \theta_m^4 (1 + |1 - \theta_m^4|) \right). \quad (124)$$

Step 4: We complete the proof by obtaining a contradiction. Lemma F.1 part 4 implies that there exists $\nu_0 > 0$ and $C > 0$ such that $A(\nu) \geq C\sqrt{\nu}$ for all $0 < \nu \leq \nu_0$. By Theorem IV.1 part 4, we can choose ν_0 small such that $\gamma(\nu) > \frac{3}{C} \left(\frac{1}{2} \theta_m^4 (1 + |1 - \theta_m^4|) \right)$ in addition to $A(\nu) \geq C\sqrt{\nu}$ for all $0 < \nu \leq \nu_0$. Since by decreasing ν_0 , the statement would be weaker, we can always assume that $\nu_0 < 1$. Thus,

$$C\sqrt{\nu} \leq A(\nu) \quad (125)$$

$$\leq \frac{2\sqrt{\nu} + \nu}{\gamma(\nu)} \left(\frac{1}{2} \theta_m^4 (1 + |1 - \theta_m^4|) \right) \quad (126)$$

$$< \frac{2\sqrt{\nu} + \nu}{\frac{3}{C} \left(\frac{1}{2} \theta_m^4 (1 + |1 - \theta_m^4|) \right)} \left(\frac{1}{2} \theta_m^4 (1 + |1 - \theta_m^4|) \right) \quad (127)$$

$$\leq C\sqrt{\nu}. \quad (128)$$

□

Lemma IV.5. *There exists $\nu_0 > 0$ such that for any $\nu_0 > \nu > 0$, the support of μ_ν^* has a finite number of points.*

Proof. We proceed by contradiction. Assume that the support of μ_ν^* has infinitely many points $\{x_i\}_{i=1}^\infty$ in increasing order with probabilities $\{\alpha_i\}_{i=1}^\infty$. Since we proved that in any bounded interval, we can only have a finite number of points, $\lim_{i \rightarrow \infty} x_i = \infty$. Note that for any $j \geq 1$, we have

$$(w_{Y|X} \circ \mu_\nu^*)(y) = \sum_{i=1}^{\infty} \alpha_i p_{x_i}(y) \quad (129)$$

$$\geq \alpha_j p_{x_j}(y), \quad (130)$$

and

$$(w_{Z|X} \circ \mu_\nu^*)(z) \geq \alpha_j q_{x_j}(z). \quad (131)$$

Therefore, for all $j \geq 1$, we can upper-bound $\phi_\nu(x)$ defined in (89) as

$$\phi_\nu(x) = \int_0^\infty p_x(y) \log \frac{p_x(y)}{(w_{Y|X} \circ \mu_\nu^*)(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{(w_{Z|X} \circ \mu_\nu^*)(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu \quad (132)$$

$$\leq \int_0^\infty p_x(y) \log \frac{p_x(y)}{\alpha_j p_{x_j}(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{\alpha_j q_{x_j}(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu \quad (133)$$

$$= \log(\theta_m^2 x + 1) - 1 - \log \frac{\alpha_j}{1 + \theta_m^2 x_j} + \frac{1 + \theta_m^2 x}{1 + \theta_m^2 x_j} - \gamma(\nu) \left(1 + x - \log \frac{\alpha_j}{1 + x_j} + \frac{1 + x}{1 + x_j} \right) - A(\nu) + \gamma(\nu)\nu \quad (134)$$

$$= \kappa + \log(\theta_m^2 x + 1) + \left(-\gamma(\nu) + \frac{\gamma(\nu)}{1 + x_j} + \frac{\theta_m^2}{1 + \theta_m^2 x_j} \right) x, \quad (135)$$

where κ is a constant not depending on x . Furthermore, the KKT condition in (88) implies that (135) is non-negative for all x_i , and since x_i can be large enough, we should have

$$-\gamma(\nu) + \frac{\gamma(\nu)}{1 + x_j} + \frac{\theta_m^2}{1 + \theta_m^2 x_j} \geq 0. \quad (136)$$

Because x_j can be large enough, we have $-\gamma(\nu) \geq 0$. This cannot be true for small ν since $\lim_{\nu \rightarrow 0^+} \gamma(\nu) = \infty$ by Theorem IV.1. \square

Lemma IV.6. *There exists $\nu_0 > 0$ such that for all $\nu_0 > \nu > 0$, μ_ν^* has a mass point at 0.*

The proof of Lemma IV.6 will require the following technical result which is a modification of [18, Lemma 1].

Lemma IV.7. *Let $f(z)$ be a PDF with mean m and $g(z)$ be a strictly monotonically increasing function, then $\int (z - m)f(z)g(z)dz > 0$.*

Proof. $(z - m)(g(z) - g(m))$ is always positive as either the product of two negative terms if $z < m$ or two positive terms if $z > m$. Thus, $(z - m)g(z) > (z - m)g(m)$ and $\int (z - m)g(z)f(z)dz > \int (z - m)g(m)f(z)dz = 0$. \square

Proof of Lemma IV.6. Let ν_0 be as in Lemma IV.5 so that μ_ν^* has finite number of mass points for all $0 < \nu \leq \nu_0$. For the sake of a contradiction, assume that μ_ν^* is a discrete probability measure over \mathcal{X} with k mass points $0 < x_1 < \dots < x_k$ with corresponding probabilities $\alpha_1, \dots, \alpha_k$. In [18], it is proved that reducing x_1 increases the mutual information $I(\mu, w_{Y|X})$. Therefore, to complete the proof, it is enough to show that $\frac{\partial \mathbb{D}(w_{Z|X} \circ \mu \| q_0)}{\partial x_1} > 0$. Defining $f(x_1, z) \triangleq (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)}$, we have

$$\frac{\partial \mathbb{D}(w_{Z|X} \circ \mu \| q_0)}{\partial x_1} = \frac{\partial}{\partial x_1} \int_{\mathcal{Z}} f(x_1, z) dz. \quad (137)$$

By Lemma C.4, $\int_{\mathcal{Z}} |f(x_1, z)| dz < \infty$, and we have

$$\frac{\partial f}{\partial x_1}(x_1, z) = \frac{\alpha_1}{(1 + x_1)^2} q_{x_1}(z) (z - \mathbb{E}_{q_{x_1}}(Z)) \left(\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} + 1 \right), \quad (138)$$

which satisfies that

$$\left| \frac{\partial f}{\partial x_1}(x_1, z) \right| \leq e^{-\frac{z}{1+x_1}} (z + x_1 + 1) (2z + \mathbb{E}_\mu(X) + 1). \quad (139)$$

The right hand side of (139), is bounded with an integrable function of z independent of x_1 , if x_1 is bounded. Hence, Theorem A.1 implies that

$$\frac{\partial \mathbb{D}(w_{Z|X} \circ \mu \| q_0)}{\partial x_1} = \alpha_1 \frac{1}{(1 + x_1)^2} \int_0^\infty (z - \mathbb{E}_{q_{x_1}}(Z)) q_{x_1}(z) \left(\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} + 1 \right) dz. \quad (140)$$

Note that

$$\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} = \log \frac{\sum_{i=1}^k \alpha_i \frac{1}{x_i + 1} e^{-\frac{z}{x_i + 1}}}{e^{-z}} \quad (141)$$

$$= \log \sum_{i=1}^k \alpha_i \frac{1}{x_i + 1} e^{z \frac{x_i}{x_i + 1}}. \quad (142)$$

Since $1 > \frac{1}{x_1 + 1} > \dots > \frac{1}{x_k + 1}$, $\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} + 1$ is strictly monotonically increasing in z . Using Lemma IV.7, $\frac{\partial \mathbb{D}(w_{Z|X} \circ \mu \| q_0)}{\partial x_1} > 0$, and hence, by decreasing x_1 , the constraint $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ still holds and $I(\mu, w_{Y|X})$ is increased. This contradicts with the definition of μ_ν^* , and therefore, there exists a mass point at zero. \square

3) *Step three: an amplitude constraint:* For a probability measure μ on \mathcal{X} and $a > 0$, we define $\mathbb{C}_a[\mu]$ as a new probability measure $\tilde{\mu}$ on \mathcal{X} such that

$$\tilde{\mu}(\cdot - \infty, x] = \begin{cases} \mu(\cdot - \infty, x] & x < a, \\ 1 & x \geq a. \end{cases} \quad (143)$$

Intuitively, $\tilde{\mu}$ is obtained by moving all probability of $]a, \infty[$ in μ to a mass point at a .

Theorem IV.2. *Let $\{\nu_n\}_{n \geq 1}$ be $o(1)$. For all $a > 1$, if n is large enough, we have $\mathbb{C}_a[\mu_{\nu_n}^*] \in \Omega_a(\nu_n)$ and*

$$\liminf_{n \rightarrow \infty} \frac{I(\mu_{\nu_n}^*, w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_{\nu_n}^* \| q_0)}} \leq \liminf_{n \rightarrow \infty} \frac{I(\mathbb{C}_a[\mu_{\nu_n}^*], w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mathbb{C}_a[\mu_{\nu_n}^*] \| q_0)}}. \quad (144)$$

To prove this result, we need the following lemmas.

Lemma IV.8. *If μ is a discrete probability measure on \mathcal{X} with finite number of mass points $x_1 < \dots < x_k$ and corresponding probabilities $\alpha_1, \dots, \alpha_k$, then*

$$\mathbb{D}(w_{Z|X} \circ \mathbb{C}_a[\mu] \| q_0) \leq \mathbb{D}(w_{Z|X} \circ \mu \| q_0), \quad (145)$$

$$I(\mathbb{C}_a[\mu], w_{Y|X}) \geq I(\mu, w_{Y|X}) - \theta_m^2 \max(\text{support}(\mu)) \mu(]a, \infty[). \quad (146)$$

Proof. Similar to (140), for all $i \in \llbracket 1, k \rrbracket$, we have

$$\frac{\partial}{\partial x_i} \mathbb{D}(w_{Z|X} \circ \mu \| q_0) = \alpha_i \frac{1}{(1+x_i)^2} \int_0^\infty (z - \mathbb{E}_{q_{x_i}}(Z)) q_{x_i}(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \geq 0. \quad (147)$$

Hence, by moving all mass points located in $]a, \infty[$ to a to obtain $\mathbb{C}_a[\mu]$, we decrease the relative entropy. Applying the same argument to the channel $w_{Y|X}$, we have $\mathbb{D}(w_{Y|X} \circ \mathbb{C}_a[\mu] \| p_0) \leq \mathbb{D}(w_{Y|X} \circ \mu \| p_0)$. Additionally, we have

$$I(\mu, w_{Y|X}) = \sum_{i=1}^k \alpha_i \mathbb{D}(p_{x_i} \| p_0) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0) \quad (148)$$

$$= \sum_{i=1}^k \alpha_i (\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0), \quad (149)$$

which implies that

$$I(\mu, w_{Y|X}) - I(\mathbb{C}_a[\mu], w_{Y|X}) \quad (150)$$

$$= \left(\sum_{i: x_i > a} \alpha_i (\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)) - \mu(]a, \infty[) (\theta_m^2 a - \log(1 + \theta_m^2 a)) \right) + (-\mathbb{D}(w_{Y|X} \circ \mu \| p_0) + \mathbb{D}(w_{Y|X} \circ \mathbb{C}_a[\mu] \| p_0)) \quad (151)$$

$$\leq \sum_{i: x_i > a} \alpha_i (\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)) - \mu(]a, \infty[) (\theta_m^2 a - \log(1 + \theta_m^2 a)) \quad (152)$$

$$\leq \sum_{i: x_i > a} \alpha_i (\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)) \quad (153)$$

$$\leq \theta_m^2 \max(\text{support}(\mu)) \mu(]a, \infty[). \quad (154)$$

□

Lemma IV.9. *For all $a > 0$, there exist $\nu_0 > 0$, $\tilde{x} \in \mathcal{X}$, and $\xi > 0$ such that for all $0 < \nu \leq \nu_0$, if $\max(\text{support}(\mu_\nu^*)) \geq \tilde{x}$, then $\mu_\nu^*(]a, \infty[) \leq 2^{-\xi \max(\text{support}(\mu_\nu^*))}$.*

Proof. Fix $\nu > 0$ small enough and suppose that $\mu \triangleq \mu_\nu^*$ has mass points $x_1 < \dots < x_k$ with corresponding probabilities $\alpha_1, \dots, \alpha_k$. Let $r(y) \triangleq (w_{Y|X} \circ \mu)(z)$ and $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$. Substituting the lower-bounds

$$r(y) \geq \frac{\mu(]a, \infty[)}{1 + \theta_m^2 x_k} e^{-\frac{y}{1 + \theta_m^2 a}}, \quad \text{and} \quad f(z) \geq \frac{\mu(]a, \infty[)}{1 + x_k} e^{-\frac{z}{1+a}}, \quad (155)$$

in the KKT condition (88) for the point $x = x_k$, we obtain

$$0 \leq -\log(\theta_m^2 x_k + 1) - 1 - \gamma(\nu)(1 + x_k) - A(\nu) + \gamma(\nu)\nu - \log \frac{\mu(]a, \infty[)}{1 + \theta_m^2 x_k} + \frac{1 + \theta_m^2 x_k}{1 + \theta_m^2 a} + \gamma(\nu) \left(-\log \frac{\mu(]a, \infty[)}{1 + x_k} + \frac{1 + x_k}{1 + a} \right). \quad (156)$$

Since $\lim_{\nu \rightarrow 0^+} \gamma(\nu)\nu = 0$, for small ν , $-1 - A(\nu) + \gamma(\nu)\nu \leq 0$, and therefore, (156) implies that

$$0 \leq -\gamma(\nu)(1+x_k) \frac{a}{1+a} + \gamma(\nu) \log(1+x_k) + \frac{1+\theta_m^2 x_k}{1+\theta_m^2 a} - (1+\gamma(\nu)) \log(\mu(\cdot|a, \infty)). \quad (157)$$

Furthermore, if x_k is large enough, we have $\log(1+x_k) \leq \frac{(1+x_k)a}{4(1+a)}$, and if ν is small enough and x_k is large enough, by Theorem IV.1 part 4, we have $\frac{1+\theta_m^2 x_k}{1+\theta_m^2 a} \leq \gamma(\nu)(1+x_k) \frac{a}{4(1+a)}$. Hence, there exist $\nu_0 > 0$ and $\tilde{x} > 0$ such that if $\nu \leq \nu_0$ and $x_k \geq \tilde{x}$, we have

$$0 \leq -\frac{1}{2}\gamma(\nu)(1+x_k) \frac{a}{1+a} - (1+\gamma(\nu)) \log(\mu(\cdot|a, \infty)), \quad (158)$$

which yields that

$$\mu(\cdot|a, \infty) \leq \exp\left(-\frac{1}{2} \frac{\gamma(\nu)}{1+\gamma(\nu)} (1+x_k) \frac{a}{1+a}\right). \quad (159)$$

Since $\lim_{\nu \rightarrow 0^+} \gamma(\nu) = \infty$, there exists $\nu_0 > 0$ such that $\inf_{\nu \in]0, \nu_0]} \frac{\gamma(\nu)}{1+\gamma(\nu)} \geq \frac{1}{2}$. Hence, for $\xi \triangleq \frac{a}{4(1+a)}$ and all $0 < \nu < \nu_0$, we have $\mu(\cdot|a, \infty) \leq 2^{-\xi x_k}$. \square

We are now ready to establish the upper bound in (7) of Theorem IV.2.

Proof of Theorem IV.2. Let $x_n^* \triangleq \max(\text{support}(\mu_{\nu_n}^*))$. By Lemma IV.5, if n is large enough $\mu_{\nu_n}^*$ is a discrete probability measure with finite number of mass points, and so is $\mathbb{C}_a[\mu_{\nu_n}^*]$. By Lemma IV.8, we have $\mathbb{D}(w_{Z|X} \circ \mathbb{C}_a[\mu_{\nu_n}^*] \| q_0) \leq \mathbb{D}(\mu_{\nu_n}^* \| q_0) = \nu_n$, and

$$\frac{I(\mathbb{C}_a[\mu_{\nu_n}^*], w_{Y|X})}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mathbb{C}_a[\mu_{\nu_n}^*] \| q_0)}} \geq \frac{I(\mu_{\nu_n}^*, w_{Y|X}) - \theta_m^2 x_n^* \mu_{\nu_n}^*(\cdot|a, \infty)}{\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_{\nu_n}^* \| q_0)}}. \quad (160)$$

Therefore, it is enough to show that

$$x_n^* \mu_{\nu_n}^*(\cdot|a, \infty) = o\left(\sqrt{\mathbb{D}(w_{Z|X} \circ \mu_{\nu_n}^* \| q_0)}\right) = o(\sqrt{\nu_n}). \quad (161)$$

To do so, we consider ν_0 , \tilde{x} , and ξ from Lemma IV.9. For n large enough such that $\frac{2}{\xi} \log \frac{1}{\nu_n} > \tilde{x}$, if $x_n^* \geq \frac{2}{\xi} \log \frac{1}{\nu_n}$, then

$$x_n^* \mu_{\nu_n}^*(\cdot|a, \infty) \leq x_n^* 2^{-\xi x_n^*}, \quad (162)$$

which is less than $2^{-\frac{1}{2}\xi x_n^*}$ for large enough n . Thus, $x_n^* \geq \frac{2}{\xi} \log \frac{1}{\nu_n}$ implies that $x_n^* \mu_{\nu_n}^*(\cdot|a, \infty) \leq \frac{1}{\nu_n}$. For the other case when $x_n^* < \frac{2}{\xi} \log \frac{1}{\nu_n}$, let $\tilde{\mu}$ be a probability distribution on \mathcal{X} with two mass points at 0 and a with probabilities $1 - \mu_{\nu_n}^*(\cdot|a, \infty)$ and $\mu_{\nu_n}^*(\cdot|a, \infty)$, respectively. Then, we have

$$\nu_n = \mathbb{D}(w_{Z|X} \circ \mu_{\nu_n}^* \| q_0) \stackrel{(a)}{\geq} \mathbb{D}(w_{Z|X} \circ \tilde{\mu} \| q_0) \stackrel{(b)}{\geq} K (\mu_{\nu_n}^*(\cdot|a, \infty))^{\frac{a+1}{a}}, \quad (163)$$

where (a) follows from the same argument as in the proof of Lemma IV.8, and (b) follows from Lemma C.6 for a constant K depending on a . Therefore, we have

$$x_n^* \mu_{\nu_n}^*(\cdot|a, \infty) \leq \frac{2}{\xi} \log \frac{1}{\nu_n} \left(\frac{\nu_n}{K}\right)^{\frac{a}{a+1}}. \quad (164)$$

Since both $\frac{1}{\nu_n}$ and $\frac{2}{\xi} \log \frac{1}{\nu_n} \left(\frac{\nu_n}{K}\right)^{\frac{a}{a+1}}$ are $o(\sqrt{\nu_n})$, we have (161). \square

4) *Step four: obtaining the bound in Theorem III.1:* We first prove a lemma that relates the constraint on the relative entropy to χ_2 divergence. Let $\tilde{\Omega}^{\geq 0}$ be the set of discrete probability measures over $[0, 1[$ with finite number of mass points.

Lemma IV.10. *Let $\epsilon > 0$ be small enough and $\{\nu_n\}_{n \geq 1}$ be a sequence of real numbers such that $\lim_{n \rightarrow \infty} \nu_n = 0$ and $2\sqrt{\nu_n} + \nu_n \leq 0.5$ for all n . There exists a sequence of probability measures $\{\lambda_n\}_{n \geq 1}$ such that $\lambda_n \in \tilde{\Omega}^{\geq 0}$ and*

$$\limsup_{n \rightarrow \infty} \frac{A(\nu_n)}{\sqrt{\nu_n}} \leq \limsup_{n \rightarrow \infty} \frac{I(\lambda_n, w_{Y|X})}{\sqrt{\frac{1}{2}\chi_2(w_{Z|X} \circ \lambda_n \| q_0)}} + \epsilon. \quad (165)$$

Proof. Let $\xi > 0$ and $\zeta \triangleq \frac{6\xi}{1-6\xi}$. Define

$$\mu_n \triangleq \mu_{\nu_n}^* \quad (166)$$

$$\mu'_n \triangleq C_{1+\zeta}[\mu_n] \quad (167)$$

$$\mu''_n \triangleq C_{a_n}[\mu'_n], \quad (168)$$

where $a'_n \triangleq \inf_{a: \mu_n(\cdot|a, \infty) \leq \nu_n^{\frac{1}{2} + \xi}} a$ and $a_n \triangleq \min(1 - \zeta, a'_n)$. Let $\{\nu_n\}_{n \geq 1}$ be a sequence of real numbers such that $\lim_{n \rightarrow \infty} \nu_n = 0$ and $2\sqrt{\nu_n} + \nu_n \leq 0.5$ for all n . By construction, we have $\mu''_n([a'_n, \infty]) \geq \nu_n^{\frac{1}{2} + \xi}$ and $\mu''_n(\cdot|a'_n, \infty) \leq \nu_n^{\frac{1}{2} + \xi}$. We next use the following lemma to upper-bound $\chi_2(w_{Z|X} \circ \mu''_n \| q_0)$.

Lemma IV.11. *Let $\mu \in \tilde{\Omega}^{\geq 0}$ such that $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ and $\max(\text{support}(\mu)) \leq a < 1$. If $2\sqrt{\nu} + \nu < 1/2$ and for some $M > 0$, we have $(w_{Z|X} \circ \mu)(M)/q_0(M) \geq e$, then*

$$\frac{1}{2}\chi_2(w_{Z|X} \circ \mu \| q_0) \leq \mathbb{D}(w_{Z|X} \circ \mu \| q_0) + \frac{1}{2}(\mathbb{E}_\mu(X))^3 \int_0^M e^{z(-1 + \frac{3a}{1+a})} dz + \frac{1}{2}(\mathbb{E}_\mu(X))^2 \int_M^\infty e^{z(-1 + \frac{2a}{1+a})} dz + 2(\mathbb{E}_\mu(X))^3. \quad (169)$$

Proof. See Appendix E. \square

We first establish a lower-bound on $(w_{Z|X} \circ \mu''_n(z))/q_0(z)$ to use Lemma IV.11. Since $a_n \leq a'_n$, we have

$$\mu''_n([a_n, \infty]) \geq \mu''_n([a'_n, \infty]) \geq \nu_n^{\frac{1}{2} + \xi}, \quad (170)$$

which yields that

$$\frac{(w_{Z|X} \circ \mu''_n)(z)}{q_0(z)} \geq \nu_n^{\frac{1}{2} + \xi} \frac{e^{\frac{a_n}{1+a_n} z}}{1 + a_n}. \quad (171)$$

Choosing $M_n = \frac{1+a_n}{a_n} \left(2 + \left(\frac{1}{2} + \zeta\right) \log \frac{1}{\nu_n}\right)$, we have $(w_{Z|X} \circ \mu''_n)(M_n)/q_0(M_n) \geq e$. Therefore, Lemma IV.11 implies that

$$\frac{1}{2}\chi_2(w_{Z|X} \circ \mu''_n \| q_0) \quad (172)$$

$$\leq \mathbb{D}(w_{Z|X} \circ \mu''_n \| q_0) + \frac{1}{2}(\mathbb{E}_{\mu''_n}(X))^3 \int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz + \frac{1}{2}(\mathbb{E}_{\mu''_n}(X))^2 \int_{M_n}^\infty e^{z(-1 + \frac{2a_n}{1+a_n})} dz + 2(\mathbb{E}_{\mu''_n}(X))^3 \quad (173)$$

$$\stackrel{(a)}{\leq} \nu_n + \frac{1}{2}(\mathbb{E}_{\mu''_n}(X))^3 \int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz + \frac{1}{2}(\mathbb{E}_{\mu''_n}(X))^2 \int_{M_n}^\infty e^{z(-1 + \frac{2a_n}{1+a_n})} dz + 2(\mathbb{E}_{\mu''_n}(X))^3 \quad (174)$$

$$\stackrel{(b)}{\leq} \nu_n \left(1 + \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz + \frac{9}{2} \int_{M_n}^\infty e^{z(-1 + \frac{2a_n}{1+a_n})} dz + 27\nu_n^{\frac{1}{2}}\right), \quad (175)$$

where (a) follows since by Lemma IV.8

$$\mathbb{D}(w_{Z|X} \circ \mu''_n \| q_0) \leq \mathbb{D}(w_{Z|X} \circ \mu'_n \| q_0) \leq \mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \nu_n, \quad (176)$$

and (b) follows since by Lemma C.2, we have $\mathbb{E}_{\mu''_n}(X) \leq 2\sqrt{\nu_n} + \nu_n \leq 3\sqrt{\nu_n}$. We now show that

$$\lim_{n \rightarrow \infty} \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz = \lim_{n \rightarrow \infty} \frac{9}{2} \int_{M_n}^\infty e^{z(-1 + \frac{2a_n}{1+a_n})} dz = \lim_{n \rightarrow \infty} 27M_0\nu_n^{\frac{1}{2}} = 0. \quad (177)$$

For the first limit, we consider two cases.

If $a_n \leq 1/4$, then $-1 + \frac{3a_n}{1+a_n} \leq -2/5$ and

$$\frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz \leq \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^\infty e^{-\frac{2z}{5}} dz \quad (178)$$

$$= \frac{135}{4}\nu_n^{\frac{1}{2}}. \quad (179)$$

If $a_n \geq 1/4$, then

$$\int_0^{M_n} e^{z(-1 + \frac{3a_n}{1+a_n})} dz \leq M_n \max\left(1, e^{M_n(-1 + \frac{3a_n}{1+a_n})}\right) \quad (180)$$

$$= \frac{1+a_n}{a_n} \left(2 + \left(\frac{1}{2} + \xi\right) \log \frac{1}{\nu_n}\right) \max\left(1, e^{\frac{1+a_n}{a_n} \left(2 + \left(\frac{1}{2} + \xi\right) \log \frac{1}{\nu_n}\right) \left(-1 + \frac{3a_n}{1+a_n}\right)}\right) \quad (181)$$

$$\leq 5 \left(2 + \left(\frac{1}{2} + \xi\right) \log \frac{1}{\nu_n}\right) \max\left(1, e^{\frac{2a_n-1}{a_n} \left(2 + \left(\frac{1}{2} + \xi\right) \log \frac{1}{\nu_n}\right)}\right). \quad (182)$$

Note that

$$5 \left(2 + \left(\frac{1}{2} + \xi\right) \log \frac{1}{\nu_n}\right) = O\left(\log \frac{1}{\nu_n}\right) \quad (183)$$

and

$$\max\left(1, e^{\frac{2a_n-1}{a_n}\left(2+\left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right)}\right) = O\left(1 + \nu_n^{-\frac{2a_n-1}{a_n}\left(\frac{1}{2}+\xi\right)}\right) \quad (184)$$

For $a_n \leq 1 - \frac{6\xi}{1-6\xi}$, we have $\frac{2a_n-1}{a_n}\left(\frac{1}{2}+\xi\right) \leq \frac{1}{2} - \xi$. For the second limit in (177), note that

$$\int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz \leq \int_{M_n}^{\infty} e^{-z\frac{\zeta}{2-\zeta}} dz. \quad (185)$$

Because $\int_0^{\infty} e^{-z\frac{\xi}{2-\xi}} < \infty$ and $M_n = \frac{1+a_n}{a_n}\left(2+\left(\frac{1}{2}+\zeta\right)\log\frac{1}{\nu_n}\right)$ goes to infinity as n goes to infinity, $\lim_{n \rightarrow \infty} \int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz = 0$. The third limit in (177) follows since $\lim_{n \rightarrow \infty} \nu_n = 0$. We thus obtain (177), which together with (175) results in

$$\limsup_{n \rightarrow \infty} \frac{\frac{1}{2}\chi_2(w_{Z|X} \circ \mu_n'' \| q_0)}{\nu_n} \leq 1. \quad (186)$$

We now consider $I(\mu_n'', w_{Y|X})$ and show that it is close to $I(\mu_n, w_{Y|X}) = A(\nu_n)$.

If $a_n = 1 - \zeta$, then by a modification of Lemma IV.8

$$I(\mu_n'', w_{Z|X}) \geq I(\mu_n', w_{Z|X}) - 2\zeta\mu_n'([1 - \zeta, \infty]) \quad (187)$$

$$= I(\mu_n', w_{Z|X}) - 2\zeta\mu_n'([1 - \zeta, \infty]) \quad (188)$$

$$\geq I(\mu_n', w_{Z|X}) - 2\zeta \frac{\mathbb{E}_{\mu_n}(X)}{1 - \zeta} \quad (189)$$

$$\geq I(\mu_n', w_{Z|X}) - 6\zeta \frac{\sqrt{\nu_n}}{1 - \zeta}. \quad (190)$$

If $a_n = a_n'$, by Lemma IV.8

$$I(\mu_n'', w_{Z|X}) \geq I(\mu_n', w_{Z|X}) - 2(1 + \zeta)\mu_n'([a_n', \infty]) \quad (191)$$

$$= I(\mu_n', w_{Z|X}) - 2(1 + \zeta)\nu_n^{\frac{1}{2} + \zeta}. \quad (192)$$

Therefore,

$$I(\mu_n'', w_{Z|X}) \geq I(\mu_n', w_{Z|X}) - \max\left(6\zeta \frac{\sqrt{\nu_n}}{1 - \zeta}, 2(1 + \zeta)\nu_n^{\frac{1}{2} + \zeta}\right) \quad (193)$$

$$\stackrel{(a)}{\geq} I(\mu_n, w_{Z|X}) - \max\left(6\zeta \frac{\sqrt{\nu_n}}{1 - \zeta}, 2(1 + \zeta)\nu_n^{\frac{1}{2} + \zeta}\right) - o(\nu_n^{\frac{1}{2}}), \quad (194)$$

where (a) follows from the argument of Theorem IV.2. Taking $\lambda_n = \mu_n'' \in \tilde{\Omega}^{\geq 0}$, by (194) and (186), we have (165) for $\epsilon = \frac{6\zeta}{1-\zeta}$. \square

Let $\mu \in \tilde{\Omega}^{\geq 0}$. We claim that

$$\frac{I(\mu, w_{Y|X})}{\sqrt{\chi_2(w_{Z|X} \circ \mu \| q_0)}} \leq \sup_{\tilde{\mu} \in \tilde{\Omega}^{\geq 0}} \frac{\mathbb{E}_{\tilde{\mu}}(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\tilde{\mu} \otimes \tilde{\mu}}\left(\frac{X_1 X_2}{1 - X_1 X_2}\right)}}. \quad (195)$$

Let us define $\tilde{\mu}$ as

$$\tilde{\mu}(A) \triangleq \frac{\mu(A \cap]0, 1])}{\mu(]0, 1])}. \quad (196)$$

In other words, $\tilde{\mu}$ is the probability measure μ conditioned to the event $]0, 1[$. We have

$$I(\mu, w_{Y|X}) \leq \mathbb{E}_{\mu}(\theta_m^2 X - \log(1 + \theta_m^2 X)) \stackrel{(a)}{=} \mu(]0, 1])\mathbb{E}_{\tilde{\mu}}(\theta_m^2 X - \log(1 + \theta_m^2 X)), \quad (197)$$

where (a) follows since $\theta_m^2 x - \log(1 + \theta_m^2 x) = 0$ for $x = 0$. Moreover,

$$\chi_2(w_{Z|X} \circ \mu \| q_0) = \mathbb{E}_{\mu \circ \mu}\left(\frac{X_1 X_2}{1 - X_1 X_2}\right) = \mu(]0, 1])^2 \mathbb{E}_{\tilde{\mu} \circ \tilde{\mu}}\left(\frac{X_1 X_2}{1 - X_1 X_2}\right) \quad (198)$$

Therefore,

$$\frac{I(\mu, w_{Y|X})}{\sqrt{\frac{1}{2}\chi_2(w_{Z|X} \circ \mu \| q_0)}} \leq \sqrt{2} \frac{\mathbb{E}_{\tilde{\mu}}(\theta_m^2 X - \log(1 + \theta_m^2 X))}{\sqrt{\mathbb{E}_{\tilde{\mu} \otimes \tilde{\mu}}\left(\frac{X_1 X_2}{1 - X_1 X_2}\right)}}. \quad (199)$$

Furthermore, with the help of Lemma F.1, Eq. (394), we have that

$$\limsup_{\nu \rightarrow 0^+} \frac{A(\nu)}{\sqrt{\nu}} \leq \sqrt{2}\theta_m^2. \quad (200)$$

Therefore, we obtain the upper-bound in (7).

V. CONCLUSION

For covert communications over non-coherent wireless channels, we showed that discrete constellations with an amplitude constraint are optimal. This differs from the results for coherent Gaussian channels in which using the phase is required to achieve the covert capacity. Supported by numerical results, we also conjectured that the optimal number of points is two and that their positions are fixed.

APPENDIX A LEIBNIZ INTEGRAL RULE

For a reader's convenience, we recall Leibniz integral rule here as it is used extensively throughout the paper.

Theorem A.1. *Let \mathcal{O} be an open subset of \mathbb{R} and Ω be a measure space. Suppose $f : \mathcal{O} \times \Omega \rightarrow \mathbb{R}$ satisfies the following conditions*

- 1) $f(x, \omega)$ is a Lebesgue-integrable function of ω for each $x \in \mathcal{O}$
- 2) For almost all $\omega \in \Omega$, the derivative $\frac{\partial f}{\partial x}$ exists for all $x \in \mathcal{O}$
- 3) There is an integrable function $\theta : \Omega \rightarrow \mathbb{R}$ such that $\left| \frac{\partial f}{\partial x}(x, \omega) \right| \leq \theta(\omega)$ for all $x \in \mathcal{O}$ and almost every $\omega \in \Omega$.

Then, for all $x \in \mathcal{O}$, we have

$$\frac{d}{dx} \int f(x, \omega) d\omega = \int \frac{\partial f}{\partial x}(x, \omega) d\omega \quad (201)$$

APPENDIX B AN ANALYTICITY CRITERION

Theorem B.1. *Let $g : \mathcal{D} \times \mathbb{R} \rightarrow \mathbb{C}$ be a function such that \mathcal{D} is a simple connected subset of \mathbb{C} , $g(\cdot, y)$ is analytic for all $y \in \mathbb{R}$, and $\sup_{z \in \mathcal{C}} \int_{\mathbb{R}} |g(z, y)| dy < \infty$ for all compact $\mathcal{C} \subset \mathcal{D}$. The function $f : z \mapsto \int_{\mathbb{R}} g(z, y) dy$ is analytic over the domain \mathcal{D} .*

Proof. The proof is a straightforward application of Fubini's theorem and Morera's theorem. Fixing any closed piecewise C^1 curve γ in \mathcal{D} , we have

$$\int_{\gamma} f(z) dz = \int_{\gamma} \int_{\mathbb{R}} g(z, y) dy dz \quad (202)$$

$$\stackrel{(a)}{=} \int_{\mathbb{R}} \int_{\gamma} g(z, y) dz dy \quad (203)$$

$$\stackrel{(b)}{=} 0, \quad (204)$$

where (a) follows from Fubini's theorem and our assumption on g , and (b) follows since $g(\cdot, z)$ is analytic and from Cauchy's integral theorem. Therefore, f satisfies the condition of Morera's theorem and is analytic. \square

APPENDIX C AUXILIARY RESULTS

We gather here essential technical tools to prove the achievability and converse results. To begin with, we bound the PDF of the output distributions of the channels $w_{Y|X}$ and $w_{Z|X}$ for an arbitrary input distribution μ .

Proposition C.1. *For any probability measure μ on \mathcal{X} with $\mathbb{E}_{\mu}(X) < \infty$ and all $y \in \mathcal{Y}, z \in \mathcal{Z}$, we have*

$$-\theta_m^2 \mathbb{E}_{\mu}(X) - y \leq \log((w_{Y|X} \circ \mu)(y)) \leq 0, \quad (205)$$

$$-\mathbb{E}_{\mu}(X) - z \leq \log((w_{Z|X} \circ \mu)(z)) \leq 0, \quad (206)$$

$$\mathbb{E}_{w_{Y|X} \circ \mu}(Y) = 1 + \theta_m^2 \mathbb{E}_{\mu}(X), \quad (207)$$

$$\mathbb{E}_{w_{Z|X} \circ \mu}(Z) = 1 + \mathbb{E}_{\mu}(X). \quad (208)$$

Proof. We only prove (205) and (207), from which (206) and (208) follow by setting $\theta_m = 1$. To obtain (205), observe that for any $x \in \mathcal{X}$, we have $p_x(y) \triangleq \frac{1}{1+\theta_m^2 x} e^{-\frac{y}{1+\theta_m^2 x}} \leq 1$, and

$$\log((w_{Y|X} \circ \mu)(y)) = \log(\mathbb{E}_\mu(p_X(y))) \quad (209)$$

$$\stackrel{(a)}{\geq} \mathbb{E}_\mu(\log(p_X(y))) \quad (210)$$

$$= \mathbb{E}_\mu\left(-\log(1 + \theta_m^2 X) - \frac{y}{1 + \theta_m^2 X}\right) \quad (211)$$

$$\stackrel{(b)}{\geq} \mathbb{E}_\mu\left(-\theta_m^2 X - \frac{y}{1 + \theta_m^2 X}\right) \quad (212)$$

$$\stackrel{(c)}{\geq} \mathbb{E}_\mu(-\theta_m^2 X - y) \quad (213)$$

$$= -\theta_m^2 \mathbb{E}_\mu(X) - y, \quad (214)$$

where (a) follows from Jensen's inequality, (b) follows from $\log(1+x) \leq x$ for $x > -1$, and (c) follows from $\mathbb{P}_\mu(X \geq 0) = 1$. To obtain (207), note that

$$\mathbb{E}_{w_{Y|X} \circ \mu}(Y) = \int_0^\infty y(w_{Y|X} \circ \mu)(y) dy \quad (215)$$

$$= \int_0^\infty y \left(\int_{\mathcal{X}} p_x(y) d\mu \right) dy \quad (216)$$

$$\stackrel{(a)}{=} \int_{\mathcal{X}} \left(\int_0^\infty y p_x(y) dy \right) d\mu \quad (217)$$

$$= \int_{\mathcal{X}} (1 + \theta_m^2 x) d\mu \quad (218)$$

$$= 1 + \theta_m^2 \mathbb{E}_\mu(X), \quad (219)$$

where (a) follows from Fubini's theorem and the fact that for all x, y , $yp_x(y) \geq 0$. \square

Lemma C.1. *Let μ be a probability measure over \mathcal{X} . If $\mathbb{D}(w_{Z|X} \circ \mu \| q_0)$ exists and is finite, then $\mathbb{E}_\mu(X) < \infty$.*

Proof. We proceed by contradiction. Consider a positive real number γ_1 and let $2\epsilon \triangleq \mu([\gamma_1, \infty))$. We have $\epsilon > 0$, because otherwise $\mathbb{E}_\mu(X) \leq \gamma_1 < \infty$. By the continuity of a probability, we have

$$\lim_{\gamma \rightarrow \infty} \mu([\gamma_1, \gamma]) = 2\epsilon. \quad (220)$$

Therefore, there exists $\gamma_2 \geq \gamma_1$ such that $\mu([\gamma_1, \gamma_2]) \geq \epsilon$. We then have

$$(w_{Z|X} \circ \mu)(z) \geq \frac{\epsilon e^{-\frac{z}{1+\gamma_1}}}{1 + \gamma_2}. \quad (221)$$

This implies that $(w_{Z|X} \circ \mu)(z) \geq q_0(z) = e^{-z}$ for all $z \geq z_0 \triangleq \frac{1+\gamma_1}{\gamma_1} \log \frac{1+\gamma_2}{\epsilon} > 0$. Since $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) < \infty$, we have

$$\infty > \int_{z_0}^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \quad (222)$$

$$\geq \int_{z_0}^\infty (w_{Z|X} \circ \mu)(z) \log \frac{\epsilon e^{-\frac{z}{1+\gamma_1}}}{e^{-z}} dz \quad (223)$$

$$\geq \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} \int_{z_0}^\infty (w_{Z|X} \circ \mu)(z) z dz \quad (224)$$

$$= \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} \int_{z_0}^\infty \int_{\mathcal{X}} \frac{e^{-\frac{z}{1+x}}}{1+x} d\mu z dz \quad (225)$$

$$= \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} \int_{\mathcal{X}} \int_{z_0}^\infty \frac{e^{-\frac{z}{1+x}}}{1+x} z dz d\mu \quad (226)$$

$$= \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} \int_{\mathcal{X}} (1+x) \left(1 + \frac{z_0}{1+x}\right) e^{-\frac{z_0}{1+x}} d\mu \quad (227)$$

$$\geq \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} \int_{\mathcal{X}} (1+x) e^{-z_0} d\mu \quad (228)$$

$$\geq \log \frac{\epsilon}{1 + \gamma_2} + \frac{\gamma_1}{1 + \gamma_1} (\mathbb{E}_\mu(X) + 1) e^{-z_0}, \quad (229)$$

which implies that $\mathbb{E}_\mu(X) < \infty$. \square

The next result shows that an upper-bound on $\mathbb{D}(w_{Z|X} \circ \mu \| q_0)$ leads to an upper-bound on $\mathbb{E}_\mu(X)$.

Lemma C.2. *For any $\nu > 0$ and for any probability measure μ on \mathcal{X} , $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ implies that $\mathbb{E}_\mu(X) \leq 2\sqrt{\nu} + \nu$.*

Proof. For any $x \in \mathbb{R}^+$, we first consider the relative entropy $\mathbb{D}(w_{Z|X} \circ \mu \| q_x)$ and show that it exists. By (206) in Proposition C.1 applied to a distribution with a single mass point at x , $|\log q_x(z)| \leq x+z$. We thus have $\int_0^\infty (w_{Z|X} \circ \mu)(z) |\log q_x(z)| dz \leq x + \mathbb{E}_{w_{Z|X} \circ \mu}(Z) = x + 1 + \mathbb{E}_\mu(X)$, which is finite by Lemma C.1. Consequently, $\int_0^\infty (w_{Z|X} \circ \mu)(z) \log q_x(z) dz$ is finite, and therefore by [26, Lemma 8.3.1], the relative entropy $\mathbb{D}(w_{Z|X} \circ \mu \| q_x)$ exists and is finite. Accordingly, we have

$$0 \geq - \int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_x(z)} dz \quad (230)$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) \left(-\log((w_{Z|X} \circ \mu)(z)) - \log(1+x) - \frac{z}{1+x} \right) dz. \quad (231)$$

Furthermore, by our assumption that $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$, we have

$$\nu \geq \int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \quad (232)$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) (\log((w_{Z|X} \circ \mu)(z)) + z) dz. \quad (233)$$

Adding the inequalities in (231) and (233), we obtain

$$\nu \geq \int_0^\infty (w_{Z|X} \circ \mu)(z) \left(-\log(1+x) + \frac{xz}{1+x} \right) dz \quad (234)$$

$$= -\log(1+x) + \frac{x}{1+x} \mathbb{E}_{w_{Z|X} \circ \mu}(Z) \quad (235)$$

$$\stackrel{(a)}{=} -\log(1+x) + \frac{x}{1+x} (\mathbb{E}_\mu(X) + 1), \quad (236)$$

where (a) follows from (208). Hence, we have

$$\mathbb{E}_\mu(X) \leq (\nu + \log(1+x)) \frac{1+x}{x} - 1 \quad (237)$$

$$\leq (\nu + x) \frac{1+x}{x} - 1. \quad (238)$$

Choosing $x = \sqrt{\nu}$, we obtain the desired upper-bound. \square

Lemma C.3. *For any probability measure μ on \mathcal{X} with $\mathbb{E}_\mu(X) < \infty$, $I(\mu, w_{Y|X})$ is well-defined and finite, and*

$$I(\mu, w_{Y|X}) = - \int_0^\infty (w_{Y|X} \circ \mu)(y) \log((w_{Y|X} \circ \mu)(y)) dy - \mathbb{E}_\mu(\log(1 + \theta_m^2 X)) - 1. \quad (239)$$

Proof. To check that $I(\mu, w_{Y|X})$ is well-defined and finite, it is enough to show that $\int \left| \log \frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right| d(w_{Y|X} \times \mu) < \infty$, which holds since

$$\int \left| \log \frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right| d(w_{Y|X} \times \mu) \leq \int (|\log p_x(y)| + |\log((w_{Y|X} \circ \mu)(y))|) d(w_{Y|X} \times \mu) \quad (240)$$

$$\stackrel{(a)}{\leq} \int (\theta_m^2(x + \mathbb{E}_\mu(X)) + 2y) d(w_{Y|X} \times \mu) \quad (241)$$

$$= 2\theta_m^2 \mathbb{E}_\mu(X) + 2\mathbb{E}_{w_{Y|X} \circ \mu}(Y) \quad (242)$$

$$\stackrel{(b)}{=} 4\theta_m^2 \mathbb{E}_\mu(X) + 2 < \infty, \quad (243)$$

where (a) follows from (205), and (b) follows from (207). Note next that

$$I(\mu, w_{Y|X}) = \mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \quad (244)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu} \left(-\log(1 + \theta_m^2 X) - \frac{Y}{1 + \theta_m^2 X} - \log((w_{Y|X} \circ \mu)(Y)) \right). \quad (245)$$

Moreover, $\mathbb{E}(\log(1 + \theta_m^2 X)) \leq \theta_m^2 \mathbb{E}(X) < \infty$ and $\mathbb{E}\left(\frac{Y}{1 + \theta_m^2 X}\right) \leq \mathbb{E}(Y) < \infty$, and therefore, we can use the linearity of expectation to write

$$\mathbb{E}_{w_{Y|X} \times \mu} \left(-\log(1 + \theta_m^2 X) - \frac{Y}{1 + \theta_m^2 X} - \log((w_{Y|X} \circ \mu)(Y)) \right) \quad (246)$$

$$= -\mathbb{E}(\log(1 + \theta_m^2 X)) - \mathbb{E}\left(\frac{Y}{1 + \theta_m^2 X}\right) - \mathbb{E}(\log((w_{Y|X} \circ \mu)(Y))) \quad (247)$$

$$= -\mathbb{E}(\log(1 + \theta_m^2 X)) - \mathbb{E}\left(\mathbb{E}\left(\frac{Y}{1 + \theta_m^2 X} \middle| X\right)\right) - \mathbb{E}(\log((w_{Y|X} \circ \mu)(Y))) \quad (248)$$

$$= -\mathbb{E}(\log(1 + \theta_m^2 X)) - \mathbb{E}\left(\frac{1 + \theta_m^2 X}{1 + \theta_m^2 X}\right) - \mathbb{E}(\log((w_{Y|X} \circ \mu)(Y))) \quad (249)$$

$$= -\mathbb{E}(\log(1 + \theta_m^2 X)) - 1 - \mathbb{E}(\log((w_{Y|X} \circ \mu)(Y))), \quad (250)$$

which completes the proof of (239). \square

Lemma C.4. *Suppose that $\mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0)$ and $\mathbb{D}(w_{Z|X} \circ \mu_2 \| q_0)$ exist and are finite for two probability measures μ_1 and μ_2 on \mathcal{X} . Then, the cross entropy $\int_0^\infty (w_{Z|X} \circ \mu_1)(z) \log(w_{Z|X} \circ \mu_2(z)) dz$ exists and is finite.*

Proof. We shall show that $\int_0^\infty (w_{Z|X} \circ \mu_1)(z) |\log((w_{Z|X} \circ \mu_2)(z))| dz < \infty$. By Lemma C.2, we know that $\mathbb{E}_{\mu_1}(X)$ and $\mathbb{E}_{\mu_2}(X)$ are finite. Therefore, we have

$$\int_0^\infty (w_{Z|X} \circ \mu_1)(z) |\log(w_{Z|X} \circ \mu_2(z))| dz \stackrel{(a)}{\leq} \int_0^\infty (w_{Z|X} \circ \mu_1)(z) (\mathbb{E}_{\mu_2}(X) + z) dz \quad (251)$$

$$= \mathbb{E}_{\mu_2}(X) + \mathbb{E}_{w_{Z|X} \circ \mu_1}(Z) \quad (252)$$

$$\stackrel{(b)}{=} \mathbb{E}_{\mu_2}(X) + 1 + \mathbb{E}_{\mu_1}(X) < \infty \quad (253)$$

where (a) follows from (206), and (b) follows from (208). \square

Lemma C.5. *Let μ be a probability measure over \mathcal{X} such that $\sup(\text{support}(\mu)) < \infty$. We then have*

$$I(\mu, w_{Y|X}) = \mathbb{E}_\mu(\theta_m^2 X - \log(1 + \theta_m^2 X)) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0). \quad (254)$$

Furthermore, if we have $\sup(\text{support}(\mu)) < 1$, then

$$\chi_2(w_{Z|X} \circ \mu \| q_0) = \mathbb{E}_{\mu \circ \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right). \quad (255)$$

Proof. We have

$$I(\mu, w_{Y|X}) = \int \log \frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} d(w_{Y|X} \times \mu) \quad (256)$$

$$= \int \log \frac{p_X(Y)}{p_0(Y)} d(w_{Y|X} \times \mu) + \int \log \frac{p_0(Y)}{(w_{Y|X} \circ \mu)(Y)} d(w_{Y|X} \times \mu) \quad (257)$$

$$= \mathbb{E}_\mu(\mathbb{D}(p_X \| p_0)) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0) \quad (258)$$

$$\stackrel{(a)}{=} \mathbb{E}_\mu(\theta_m^2 X - \log(1 + \theta_m^2 X)) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0), \quad (259)$$

where (a) follows from the straightforward calculation of the relative entropy between two exponential distribution. Additionally, we have

$$\chi_2(w_{Z|X} \circ \mu \| q_0) = \int_0^\infty \frac{(w_{Z|X} \circ \mu)(z)^2}{q_0(z)} dz - 1 \quad (260)$$

$$= \int_0^\infty \mathbb{E}_{\mu \otimes \mu} \left(\frac{1}{(1 + X_1)(1 + X_2)} e^{z(1 - \frac{1}{1+X_1} - \frac{1}{1+X_2})} \right) dz - 1 \quad (261)$$

$$\stackrel{(a)}{=} \mathbb{E}_{\mu \otimes \mu} \left(\int_0^\infty \frac{1}{(1 + X_1)(1 + X_2)} e^{z(1 - \frac{1}{1+X_1} - \frac{1}{1+X_2})} dz \right) - 1 \quad (262)$$

$$= \mathbb{E}_{\mu \otimes \mu} \left(\frac{1}{1 - X_1 X_2} \right) - 1 \quad (263)$$

$$= \mathbb{E}_{\mu \otimes \mu} \left(\frac{X_1 X_2}{1 - X_1 X_2} \right), \quad (264)$$

where (a) follows from Fubini theorem and $\frac{1}{(1+X_1)(1+X_2)} e^{z(1 - \frac{1}{1+X_1} - \frac{1}{1+X_2})} \geq 0$ almost surely. \square

Lemma C.6. *If $a > 1$ and $\beta > 0$ is small enough, then*

$$\mathbb{D}(\beta q_a + (1 - \beta)q_0 \| q_0) = \beta^{1+\frac{1}{a}} (1+a)^{-1-\frac{1}{a}} \left(1 + \frac{1}{a}\right) \left(\frac{\Gamma(-\frac{1}{a}) \Gamma(2 + \frac{1}{a})}{(1 + \frac{1}{a})^2} + a^2 \Gamma\left(1 - \frac{1}{a}\right) \Gamma\left(1 + \frac{1}{a}\right) \right) + O(\beta^2), \quad (265)$$

where $\Gamma(x) \triangleq \int_0^\infty y^{x-1} e^{-y} dy$.

If $a < 1$ and $\beta > 0$ is small enough, then

$$\mathbb{D}(\beta q_a + (1 - \beta)q_0 \| q_0) = \frac{a^2}{2(1 - a^2)} \beta^2 + o(\beta^2). \quad (266)$$

Proof. We only consider the case where $a > 1$ and the other case follows from similar approach. By definition, we have

$$\mathbb{D}(\beta q_a + (1 - \beta)q_0 \| q_0) = \int_0^\infty (\beta q_a(z) + (1 - \beta)q_0(z)) \log \left(\frac{\beta q_a(z) + (1 - \beta)q_0(z)}{q_0(z)} \right) dz \quad (267)$$

$$= \int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1 - \beta)e^{-z} \right) \log \left(1 - \beta + \frac{\beta}{1+a} e^{\frac{az}{1+a}} \right) dz \quad (268)$$

$$= \log(1 - \beta) + \int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1 - \beta)e^{-z} \right) \log \left(1 + \frac{\beta}{(1 - \beta)(1 + a)} e^{\frac{az}{1+a}} \right) dz. \quad (269)$$

By substitution $u \triangleq e^{\frac{az}{1+a}}$ in the above integral, we obtain

$$\begin{aligned} \int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1 - \beta)e^{-z} \right) \log \left(1 + \frac{\beta}{(1 - \beta)(1 + a)} e^{\frac{az}{1+a}} \right) dz = \\ \left(1 + \frac{1}{a}\right) \int_1^\infty \left((1 - \beta)u^{-2-\frac{1}{a}} + \frac{\beta}{1+a} u^{-1-\frac{1}{a}} \right) \log \left(1 + \frac{\beta}{(1 - \beta)(1 + a)} u \right) du \end{aligned} \quad (270)$$

Note next that for all real numbers λ_1, λ_2 , a primitive function of $u^{\lambda_1} \log(1 + \lambda_2 u)$ is

$$\int u^{\lambda_1} \log(1 + \lambda_2 u) du = \frac{u^{\lambda_1+1} ({}_2F_1(1, \lambda_1 + 1; \lambda_1 + 2; -\lambda_2 u) + (\lambda_1 + 1) \log(\lambda_2 u + 1) - 1)}{(\lambda_1 + 1)^2} + \text{constant}, \quad (271)$$

where ${}_2F_1(a, b; c; x)$ is the hypergeometric function. Additionally, for $\lambda_1 < -1$, the limit of this primitive function at $u = \infty$ is

$$\frac{\lambda_2^{-\lambda_1-1} \Gamma(2 + \lambda_1) \Gamma(-\lambda_1)}{(\lambda_1 + 1)^2}. \quad (272)$$

Therefore, if we define $\lambda \triangleq \frac{\beta}{(1 - \beta)(1 + a)}$, by linearity of integral, we have

$$\int_1^\infty \left((1 - \beta)u^{-2-\frac{1}{a}} + \frac{\beta u^{-\frac{1}{a}}}{1+a} \right) \log \left(1 + \frac{\beta}{(1 - \beta)(1 + a)} u \right) du \quad (273)$$

$$= (1 - \beta) \left(\frac{\lambda^{1+\frac{1}{a}} \Gamma(-\frac{1}{a}) \Gamma(2 + \frac{1}{a})}{(1 + \frac{1}{a})^2} - \frac{{}_2F_1(1, -1 - \frac{1}{a}; -\frac{1}{a}; -\lambda) - (1 + \frac{1}{a}) \log(\lambda + 1) - 1}{(1 + \frac{1}{a})^2} \right) + \quad (274)$$

$$\frac{\beta}{1+a} \left(\frac{\lambda^{\frac{1}{a}} \Gamma(1 - \frac{1}{a}) \Gamma(1 + \frac{1}{a})}{(\frac{1}{a})^2} - \frac{{}_2F_1(1, -\frac{1}{a}; 1 - \frac{1}{a}; -\lambda) - (\frac{1}{a}) \log(\lambda + 1) - 1}{(\frac{1}{a})^2} \right) \quad (275)$$

$$\stackrel{(a)}{=} (1 - \beta) \left(\frac{\lambda^{1+\frac{1}{a}} \Gamma(-\frac{1}{a}) \Gamma(2 + \frac{1}{a})}{(1 + \frac{1}{a})^2} - \frac{\left(1 + \frac{\lambda(1+\frac{1}{a})}{-\frac{1}{a}}\right) - (1 + \frac{1}{a})\lambda - 1 + O(\beta^2)}{(1 + \frac{1}{a})^2} \right) + \quad (276)$$

$$\frac{\beta}{1+a} \left(\frac{\lambda^{\frac{1}{a}} \Gamma(1 - \frac{1}{a}) \Gamma(1 + \frac{1}{a})}{(\frac{1}{a})^2} - \frac{\left(1 + \frac{\lambda^{\frac{1}{a}}}{1 - \frac{1}{a}}\right) - (\frac{1}{a})\lambda + O(\beta^2) - 1}{(\frac{1}{a})^2} \right) \quad (277)$$

$$= (1 - \beta) \left(\frac{\lambda^{-1-\frac{1}{a}} \Gamma(-\frac{1}{a}) \Gamma(2 + \frac{1}{a})}{(1 + \frac{1}{a})^2} - \frac{\frac{\lambda(1+\frac{1}{a})}{-\frac{1}{a}} - (1 + \frac{1}{a})\lambda + O(\beta^2)}{(1 + \frac{1}{a})^2} \right) + \quad (278)$$

$$\lambda(1-\beta) \left(\frac{\lambda^{-\frac{1}{a}} \Gamma(1-\frac{1}{a}) \Gamma(1+\frac{1}{a})}{(\frac{1}{a})^2} - \frac{\frac{\lambda^{\frac{1}{a}}}{1-\frac{1}{a}} - (\frac{1}{a})\lambda + O(\beta^2)}{(\frac{1}{a})^2} \right), \quad (279)$$

where (a) follows since for x going to zero ${}_2F_1(a, b; c; x) = 1 + abx/c + O(x^2)$ and $\log(1+x) = x + O(x^2)$ by Taylor's expansion. By rearranging the terms in above expression and disregarding the higher order terms, we obtain

$$\lambda^{1+\frac{1}{a}}(1-\beta) \left(\frac{\Gamma(-\frac{1}{a}) \Gamma(2+\frac{1}{a})}{(1+\frac{1}{a})^2} + \frac{\Gamma(1-\frac{1}{a}) \Gamma(1+\frac{1}{a})}{(\frac{1}{a})^2} \right) + \lambda \frac{1+a}{1+\frac{1}{a}}(1-\beta) + O(\beta^2) \quad (280)$$

$$= \beta^{1+\frac{1}{a}} \left(\frac{1}{(1-\beta)(1+a)} \right)^{1+\frac{1}{a}} (1-\beta) \left(\frac{\Gamma(-\frac{1}{a}) \Gamma(2+\frac{1}{a})}{(1+\frac{1}{a})^2} + \frac{\Gamma(1-\frac{1}{a}) \Gamma(1+\frac{1}{a})}{(\frac{1}{a})^2} \right) + \frac{\beta}{1+\frac{1}{a}} + O(\beta^2). \quad (281)$$

Combining (269), (270), and (281), we have

$$\mathbb{D}(\beta q_a + (1-\beta)q_0 \| q_0) \quad (282)$$

$$= \beta^{1+\frac{1}{a}} \left(\frac{1}{(1-\beta)(1+a)} \right)^{1+\frac{1}{a}} (1-\beta)(1+\frac{1}{a}) \left(\frac{\Gamma(-\frac{1}{a}) \Gamma(2+\frac{1}{a})}{(1+\frac{1}{a})^2} + \frac{\Gamma(1-\frac{1}{a}) \Gamma(1+\frac{1}{a})}{(\frac{1}{a})^2} \right) + O(\beta^2). \quad (283)$$

□

APPENDIX D ERROR EXPONENTS ANALYSIS

Lemma D.1. For a probability measure on \mathcal{X} , μ , for which we have $\max(\text{support}(\mu)) \triangleq x_{\max} < \infty$ and $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ and for any $A > 0$, it holds that

$$\begin{aligned} & \mathbb{E}_{w_{Y|X} \times \mu} \left(\log^2 \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \right) \\ & \leq 2(3 + x_{\max})(2\sqrt{\nu} + \nu)(1 + \theta_m^2 x_{\max})^4 (e^{Ax_{\max}} A + \theta_m^2)^2 e^{2A} + 20((1 + \theta_m^2 x_{\max}) + A)^2 e^{-A}. \end{aligned} \quad (284)$$

Proof. We first define $f(x) \triangleq \int_0^\infty p_x(y) \log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) dy$ for which we have

$$f(x) = \int_0^A p_x(y) \log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) dy + \int_A^\infty p_x(y) \log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) dy, \quad (285)$$

for any $A > 0$. To upper-bound the first term, we note that

$$\left| \frac{(w_{Y|X} \circ \mu)(y)}{p_x(y)} - 1 \right| = \left| \frac{\mathbb{E}_\mu \left(\frac{1}{1+\theta_m^2 \tilde{X}} e^{-\frac{y}{1+\theta_m^2 \tilde{X}}} \right) - 1}{p_x(y)} \right| \quad (286)$$

$$= \left| \mathbb{E}_\mu \left(\frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} e^{\frac{y(x-\tilde{X})}{(1+\theta_m^2 \tilde{X})(1+\theta_m^2 x)}} - 1 \right) \right| \quad (287)$$

$$\leq \mathbb{E}_\mu \left(\left| \frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} e^{\frac{y(x-\tilde{X})}{(1+\theta_m^2 \tilde{X})(1+\theta_m^2 x)}} - 1 \right| \right) \quad (288)$$

$$\leq \mathbb{E}_\mu \left(\left| \frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} \left(e^{\frac{y(x-\tilde{X})}{(1+\theta_m^2 \tilde{X})(1+\theta_m^2 x)}} - 1 \right) \right| \right) + \mathbb{E}_\mu \left(\left| \frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} - 1 \right| \right). \quad (289)$$

Considering each term separately in the above expression, we have

$$\mathbb{E}_\mu \left(\left| \frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} \left(e^{\frac{y(x-\tilde{X})}{(1+\theta_m^2 \tilde{X})(1+\theta_m^2 x)}} - 1 \right) \right| \right) \leq (1 + \theta_m^2 x_{\max}) \mathbb{E}_\mu \left(\left| e^{\frac{y(x-\tilde{X})}{(1+\theta_m^2 \tilde{X})(1+\theta_m^2 x)}} - 1 \right| \right) \quad (290)$$

$$\stackrel{(a)}{\leq} (1 + \theta_m^2 x_{\max}) e^{y x_{\max}} \mathbb{E}_\mu \left(\left| \frac{y(x-\tilde{X})}{(1 + \theta_m^2 \tilde{X})(1 + \theta_m^2 x)} \right| \right) \quad (291)$$

$$\leq (1 + \theta_m^2 x_{\max}) e^{y x_{\max}} y \left(x + \mathbb{E}_\mu(\tilde{X}) \right), \quad (292)$$

where (a) follows from the mean value theorem and an upper-bound on derivative. For the next term in (289), we have

$$\mathbb{E}_\mu \left(\left| \frac{1 + \theta_m^2 x}{1 + \theta_m^2 \tilde{X}} - 1 \right| \right) = \theta_m^2 \mathbb{E}_\mu \left(\left| \frac{x - \tilde{X}}{1 + \theta_m^2 \tilde{X}} \right| \right) \quad (293)$$

$$\leq \theta_m^2 \mathbb{E}_\mu (|x - \tilde{X}|) \quad (294)$$

$$\leq \theta_m^2 (x + \mathbb{E}_\mu(\tilde{X})). \quad (295)$$

Combining these two inequalities, we obtain

$$\left| \frac{(w_{Y|X} \circ \mu)(y)}{p_x(y)} - 1 \right| \leq (x + \mathbb{E}_\mu(\tilde{X})) ((1 + \theta_m^2 x_{\max}) e^{y x_{\max}} y + \theta_m^2). \quad (296)$$

Hence, using the inequalities $\log^2(x) \leq (1-x)^2(1+x^{-2})$ for $x > -1$ and $\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \leq (1 + \theta_m^2 x_{\max}) e^y$, we have

$$\log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) \leq \left((x + \mathbb{E}_\mu(\tilde{X})) ((1 + \theta_m^2 x_{\max}) e^{y x_{\max}} y + \theta_m^2) \right)^2 \left(1 + ((1 + \theta_m^2 x_{\max}) e^y)^2 \right). \quad (297)$$

This yields that

$$\int_0^A p_x(y) \log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) dy \quad (298)$$

$$\leq \sup_{y \in [0, A]} \left((x + \mathbb{E}_\mu(\tilde{X})) ((1 + \theta_m^2 x_{\max}) e^{y x_{\max}} y + \theta_m^2) \right)^2 \left(1 + ((1 + \theta_m^2 x_{\max}) e^y)^2 \right) \quad (299)$$

$$= (x + \mathbb{E}_\mu(\tilde{X}))^2 ((1 + \theta_m^2 x_{\max}) e^{A x_{\max}} A + \theta_m^2)^2 \left(1 + ((1 + \theta_m^2 x_{\max}) e^A)^2 \right) \quad (300)$$

$$\leq 2 (x + \mathbb{E}_\mu(\tilde{X}))^2 (1 + \theta_m^2 x_{\max})^4 (e^{A x_{\max}} A + \theta_m^2)^2 e^{2A}. \quad (301)$$

For the second term in (285), if $x \leq x_{\max}$, then we have

$$\int_A^\infty p_x(y) \log^2 \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) dy \quad (302)$$

$$\leq 4 \int_A^\infty p_x(y) (\log(1 + \theta_m^2 x_{\max}) + y)^2 dy \quad (303)$$

$$= 4 \left(-(\log^2(1 + \theta_m^2 x_{\max}) + 2 \log(1 + \theta_m^2 x_{\max}) (y + 1 + \theta_m^2 x) + (y^2 + 2(1 + \theta_m^2 x)y + 2(1 + \theta_m^2 x)^2)) e^{-\frac{y}{1 + \theta_m^2 x}} \right) \Big|_A^\infty \quad (304)$$

$$= 4 (\log^2(1 + \theta_m^2 x_{\max}) + 2 \log(1 + \theta_m^2 x_{\max}) (A + 1 + \theta_m^2 x) + (A^2 + 2(1 + \theta_m^2 x)A + 2(1 + \theta_m^2 x)^2)) e^{-\frac{A}{1 + \theta_m^2 x}} \quad (305)$$

$$\leq 20 ((1 + \theta_m^2 x_{\max}) + A)^2 e^{-A}. \quad (306)$$

Therefore, for all $x \in \mathcal{X}$, it holds that

$$f(x) \leq 2 (x + \mathbb{E}_\mu(\tilde{X}))^2 (1 + \theta_m^2 x_{\max})^4 (e^{A x_{\max}} A + \theta_m^2)^2 e^{2A} + 20 ((1 + \theta_m^2 x_{\max}) + A)^2 e^{-A}, \quad (307)$$

which implies that

$$\mathbb{E}_{w_{Y|X} \times \mu} \left(\log^2 \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \right) \quad (308)$$

$$= \mathbb{E}_\mu(f(X)) \quad (309)$$

$$\leq 2 \mathbb{E}_\mu \left((X + \mathbb{E}_\mu(\tilde{X}))^2 \right) (1 + \theta_m^2 x_{\max})^4 (e^{A x_{\max}} A + \theta_m^2)^2 e^{2A} + 20 ((1 + \theta_m^2 x_{\max}) + A)^2 e^{-A}. \quad (310)$$

Finally, by Lemma C.2, $\mathbb{E}_\mu \left((X + \mathbb{E}_\mu(\tilde{X}))^2 \right) = \mathbb{E}_\mu(X^2) + 3(\mathbb{E}_\mu(X))^2 \leq (3 + x_{\max})(\nu + 2\sqrt{\nu})$ which completes the proof. \square

Proof of Lemma IV.1. We fix μ with $\sup(\text{support}(\mu)) \triangleq \tilde{x} < \infty$ and use Theorem A.1 along with induction to show that for a small neighborhood around zero and all $i \geq 0$, we have

$$\frac{\partial^i g}{\partial s^i}(s, \mu) = \mathbb{E}_{w_{Y|X} \times \mu} \left(\log^i \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right)^s \right), \quad (311)$$

where

$$g(s, \mu) \triangleq \mathbb{E}_{w_{Y|X} \times \mu} \left(\left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right)^s \right). \quad (312)$$

The statement is true for $i = 0$ by definition. For $i > 0$, we take $\mathcal{O} = [0, \tilde{s}]$, $\Omega = (\mathcal{X} \times \mathcal{Y}, w_{Y|X} \times \mu)$, and $f(s, x, y) = \log^{i-1} \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s$ and check the three conditions in Theorem A.1:

1) For $x \leq \tilde{x}$, we have

$$|f(s, x, y)| = \left| \log^{i-1} \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s \right| \quad (313)$$

$$\stackrel{(a)}{\leq} \left| (\theta_m^2 (\mathbb{E}_\mu(X) + x) + 2y)^{i-1} \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s \right| \quad (314)$$

$$\leq \left| (2\theta_m^2 \tilde{x} + 2y)^{i-1} \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s \right| \quad (315)$$

$$\leq |(2\theta_m^2 \tilde{x} + 2y)|^{i-1} (1 + \tilde{x})^s e^{\frac{s\tilde{x}y}{1+\tilde{x}}}, \quad (316)$$

where (a) follows from Proposition C.1. Because the above upper-bound does not depend on x , we can write

$$\mathbb{E}_{w_{Y|X} \times \mu} (|f(s, X, Y)|) \leq \mathbb{E}_{w_{Y|X} \circ \mu} \left(|(2\theta_m^2 \tilde{x} + 2Y)|^{i-1} (1 + \tilde{x})^s e^{\frac{s\tilde{x}Y}{1+\tilde{x}}} \right). \quad (317)$$

Moreover, note that the moment generating function of a random variable with exponential distribution and mean λ exists in $[0, \lambda)$, which implies that the moment generating function of distribution $w_{Y|X} \circ \mu$ exists in $[0, 1/(1 + \tilde{x}))$. Hence, there exists \tilde{s} depending on \tilde{x} such that

$$\mathbb{E}_{w_{Y|X} \circ \mu} \left(|(2\theta_m^2 \tilde{x} + 2Y)|^{i-1} (1 + \tilde{x})^s e^{\frac{s\tilde{x}Y}{1+\tilde{x}}} \right) < \infty. \quad (318)$$

2) Since for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, it holds that $0 < \frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} < \infty$, $\frac{\partial f}{\partial s}$ exists, and we have

$$\frac{\partial f}{\partial s}(s, x, y) = \log^i \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s. \quad (319)$$

3) Similar to the first part, we can upper-bound the partial derivative as

$$\left| \frac{\partial f}{\partial s}(s, x, y) \right| = \left| \log^i \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right) \left(\frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right)^s \right| \quad (320)$$

$$\leq |(2\theta_m^2 \tilde{x} + 2y)|^i (1 + \tilde{x})^s e^{\frac{s\tilde{x}y}{1+\tilde{x}}} \quad (321)$$

The above bound is increasing in s . Thus, by choosing \tilde{s} small enough such that the expectation is finite for $s = \tilde{s}$, we can choose

$$\theta(x, y) \triangleq |(2\theta_m^2 \tilde{x} + 2y)|^i (1 + \tilde{x})^{\tilde{s}} e^{\frac{\tilde{s}\tilde{x}y}{1+\tilde{x}}}. \quad (322)$$

Then, $\mathbb{E}_{w_{Y|X} \times \mu}(\theta(X, Y)) < \infty$ and for all $s \leq \tilde{s}$, we have $\left| \frac{\partial f}{\partial s}(s, x, y) \right| \leq \theta(x, y)$.

We can now use Theorem A.1 and obtain

$$\frac{\partial}{\partial s} \mathbb{E}_{w_{Y|X} \times \mu} \left(\log^{i-1} \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right)^s \right) \quad (323)$$

$$= \frac{\partial}{\partial s} \mathbb{E}_{w_{Y|X} \times \mu} (f(s, X, Y)) \quad (324)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu} \left(\frac{\partial}{\partial s} f(s, X, Y) \right) \quad (325)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu} \left(\log^i \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) \left(\frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right)^s \right). \quad (326)$$

Therefore, the induction hypothesis implies (311). By the chain rule, $\phi_{\text{rel}}(s, \mu)$ is also a smooth function on an interval $[0, \tilde{s}]$ for all μ with $\sup(\text{support}(\mu)) \leq \tilde{x}$. Hence, we can use Taylor's theorem to obtain

$$\phi(s, \mu) = \phi_{\text{rel}}(0, \mu) + \frac{\partial \phi_{\text{rel}}}{\partial s}(0, \mu) s + \frac{\partial^2 \phi_{\text{rel}}}{\partial s^2}(0, \mu) \frac{s^2}{2} + \frac{\partial^3 \phi_{\text{rel}}}{\partial s^3}(\eta, \mu) \frac{s^3}{6}, \quad (327)$$

for some $\eta \in [0, \tilde{s}]$. The derivatives of ϕ_{rel} would be

$$\phi_{\text{rel}}(0, \mu) = -\log(g(0, \mu)) = 0 \quad (328)$$

$$\frac{\partial \phi_{\text{rel}}}{\partial s}(0, \mu) = -\frac{\frac{\partial g}{\partial s}(0, \mu)}{g(0, \mu)} = -\mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) = -I(\mu, w_{Y|X}) \quad (329)$$

$$\frac{\partial^2 \phi_{\text{rel}}}{\partial s^2}(0, \mu) = -\frac{g(0, \mu) \frac{\partial^2 g}{\partial s^2}(0, \mu) - \left(\frac{\partial g}{\partial s}(0, \mu) \right)^2}{g(0, \mu)^2} = -\mathbb{E}_{w_{Y|X} \times \mu} \left(\log^2 \frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) + I(\mu, w_{Y|X})^2. \quad (330)$$

Moreover, Lemma D.1 yields that

$$\frac{\partial^2 \phi_{\text{rel}}}{\partial s^2}(0, \mu) \geq -2(3 + \tilde{x})(2\sqrt{\nu} + \nu)(1 + \theta_m^2 \tilde{x})^4 \left(e^{A\tilde{x}} A + \theta_m^2 \right)^2 e^{2A} + 20 \left((1 + \theta_m^2 \tilde{x}) + A \right)^2 e^{-A} \quad (331)$$

$$\geq -B_1 \left((2\sqrt{\nu} + \nu) e^{2A\tilde{x} + 2A} A^2 + A^2 e^{-A} \right), \quad (332)$$

for some B_1 depending on θ_m^2 and \tilde{x} . With similar arguments as we had to check the third condition of Theorem A.1, we can prove that there exists B_2 depending on \tilde{x} , such that for all $\eta \in [0, \tilde{s}]$, we have

$$\left| \frac{\partial^3 \phi_{\text{rel}}}{\partial s^3}(\eta, \mu) \right| \leq B_2. \quad (333)$$

Choosing $B = \max(B_1/2, B_2/6)$ completes the proof. \square

APPENDIX E PROOF OF LEMMA IV.11 AND IV.2

We first introduce some notation and facts, which will be useful in both proofs. Let $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$ and $\phi(z) \triangleq f(z)/q_0(z) - 1$. Defining P_X as the associated PMF of μ , we can write $\phi(z) = \sum_x P_X(x) \frac{e^{\frac{zx}{1+x}}}{1+x} - 1$, which is increasing and

$$\phi(z) \geq \phi(0) = \mathbb{E}_\mu \left(\frac{1}{1+X} \right) - 1 \geq -\mathbb{E}_\mu(X) \geq -2\sqrt{\nu} - \nu \geq -0.5. \quad (334)$$

Furthermore, there exists a unique M_0 such that $\phi(z) \leq 0$ if and only if $z \leq M_0$.

Proof of Lemma IV.2. Using the bound $\log(1+x) \leq x - x^2/2 + x^3/3$ for $x > -1$, we obtain

$$\mathbb{D}(w_{Z|X} \circ \mu \| q_0) = \int_0^\infty q_0(1+\phi) \log(1+\phi) \quad (335)$$

$$\leq \int_0^M q_0 \phi + \frac{1}{2} \int_0^M q_0 \phi^2 - \frac{1}{6} \int_0^M q_0 \phi^3 + \frac{1}{3} \int_0^M q_0 \phi^4 + \int_M^\infty f \log(f/q_0). \quad (336)$$

We consider each term separately.

1) We have

$$\int_0^M q_0 \phi = \int_0^M f - \int_0^M q_0 \quad (337)$$

$$= e^{-M} - \sum_x P_X(x) e^{-\frac{M}{1+x}} \quad (338)$$

$$\leq 0. \quad (339)$$

2) We have

$$\frac{1}{2} \int_0^M q_0 \phi^2 \leq \frac{1}{2} \int_0^\infty q_0 \phi^2 \quad (340)$$

$$= \frac{1}{2} \chi_2(w_{Z|X} \circ \mu \| q_0). \quad (341)$$

3) We have

$$-\frac{1}{6} \int_0^M q_0 \phi^3 \leq -\frac{1}{6} \int_0^{M_0} q_0 \phi^3 \quad (342)$$

$$\leq \frac{1}{6} M_0 (\mathbb{E}_\mu(X))^3 \quad (343)$$

$$\stackrel{(a)}{\leq} \frac{1}{3} (\mathbb{E}_\mu(X))^3, \quad (344)$$

where (a) follows since $M_0 \leq 2$ by the argument in the proof of Lemma IV.11.

4) We have

$$\int_0^M q_0 \phi^4 = \int_0^{M_0} q_0 \phi^4 + \int_{M_0}^M q_0 \phi^4 \quad (345)$$

$$\leq M_0 (\mathbb{E}_\mu(X))^4 + \int_{M_0}^M e^{-z} (\mathbb{E}_\mu(X) e^{\frac{az}{1+a}})^4 dz \quad (346)$$

$$= (\mathbb{E}_\mu(X))^4 \left(2 + \int_0^M e^{z(-1+\frac{4a}{1+a})} dz \right). \quad (347)$$

5) We have

$$\int_M^\infty f \log(f/q_0) \leq \frac{a}{1+a} \int_M^\infty f(z) z dz \quad (348)$$

$$\leq \frac{a}{1+a} \left(\int_M^\infty e^{-\frac{z}{1+\epsilon}} z dz + \frac{1}{\mathbb{E}_\mu(X) + \epsilon} \int_M^\infty e^{-\frac{z}{1+\epsilon}} dz \right). \quad (349)$$

□

Lemma IV.11. We use the notations introduced in the beginning of Appendix E. Since we have $\log(1+x) \geq x - x^2/2 + \mathbf{1}\{x \leq 0\} 2x^3/3$ for $x \geq -0.5$, we have for all $z \in \mathcal{Z}$,

$$f(z) \log(\phi(z) + 1) \geq f(z) (\phi(z) - \phi^2(z)/2 + \mathbf{1}\{\phi(z) \leq 0\} 2\phi(z)^3/3). \quad (350)$$

We therefore obtain

$$\mathbb{D}(f\|q_0) = \int_0^\infty f \log(\phi + 1) \quad (351)$$

$$\geq \int_0^M f (\phi - \phi^2/2) + \int_0^{M_0} 2f\phi^3/3 + \int_M^\infty f \log(f/q_0). \quad (352)$$

We consider each term separately in the following.

1) We have

$$\int_0^M f (\phi - \phi^2/2) = \int_0^M q_0 (\phi + 1) (\phi - \phi^2/2) \quad (353)$$

$$= \int_0^M q_0 \phi^2/2 + \int_0^M q_0 \phi - \int_0^M q_0 \phi^3/2 \quad (354)$$

$$= \int_0^\infty q_0 \phi^2/2 + \int_0^M q_0 \phi - \int_0^M q_0 \phi^3/2 - \int_M^\infty q_0 \phi^2/2. \quad (355)$$

We again separately lower-bound each term in the above expression.

a) We have by definition,

$$\int_0^\infty q_0 \phi^2/2 = \frac{1}{2} \chi_2(f\|q_0). \quad (356)$$

b) We have

$$\int_0^M q_0 \phi = \int_0^M (f - q_0) \quad (357)$$

$$\geq - \int_M^\infty f. \quad (358)$$

c) To lower-bound $-\int_0^M q_0 \phi^3/2$, we first upper-bound ϕ as follows.

$$\phi(z) = \sum_x P_X(x) \frac{e^{\frac{xz}{1+x}}}{1+x} - 1 \quad (359)$$

$$\leq \sum_x P_X(x) e^{\frac{xz}{1+a}} - 1 \quad (360)$$

$$\stackrel{(a)}{\leq} \sum_x P_X(x) (1 + (e^{\frac{az}{1+a}} - 1)x) - 1 \quad (361)$$

$$= (e^{\frac{az}{1+a}} - 1) \mathbb{E}_\mu(X) \quad (362)$$

$$= e^{\frac{az}{1+a}} \mathbb{E}_\mu(X), \quad (363)$$

where (a) follows since $e^{\frac{xz}{1+a}} \leq 1 + (e^{\frac{az}{1+a}} - 1)x$ for $x \in [0, a]$. Since $q_0 > 0$ and $x \mapsto x^3$ is increasing, we have

$$\int_0^M q_0 \phi^3 \leq \int_0^M e^{-z} (e^{\frac{az}{1+a}} \mathbb{E}_\mu(X))^3 dz \quad (364)$$

$$= (\mathbb{E}_\mu(X))^3 \int_0^M e^{z(-1+\frac{3a}{1+a})} dz. \quad (365)$$

d) Since $\phi(z) \geq 0$ for $z \geq M \geq M_0$ and $x \mapsto x^2$ is increasing for $x \geq 0$, we have

$$\int_M^\infty q_0 \phi^2 \leq \int_M^\infty e^{-z} (e^{\frac{az}{1+a}} \mathbb{E}_\mu(X))^2 dz \quad (366)$$

$$= (\mathbb{E}_\mu(X))^2 \int_0^M e^{z(-1+\frac{2a}{1+a})} dz. \quad (367)$$

As a conclusion, we obtain that

$$\int_0^M f(\phi - \phi^2/2) \geq \frac{1}{2} \chi_2(f \| q_0) - \frac{1}{2} (\mathbb{E}_\mu(X))^2 \int_0^M e^{z(-1+\frac{2a}{1+a})} dz - \frac{1}{2} (\mathbb{E}_\mu(X))^3 \int_0^M e^{z(-1+\frac{3a}{1+a})} dz - \int_M^\infty f. \quad (368)$$

2) Using $|f(z)| \leq 1$ for all $z \in \mathcal{Z}$ and $0 \geq \phi(z) \geq -\mathbb{E}_\mu(X)$ for all $0 \leq z \leq M_0$, we have

$$\int_0^{M_0} f \phi^3 \geq -M_0 (\mathbb{E}_\mu(X))^3. \quad (369)$$

We now show that $M_0 \leq 2$, for which it is enough to show that $f(2) \geq q_0(2)$. Note that

$$\log f(2) \stackrel{(a)}{\geq} \sum_x P_X(x) \log q_x(2) \quad (370)$$

$$= -\mathbb{E}_\mu(\log(1+X)) - 2\mathbb{E}_\mu\left(\frac{1}{1+X}\right) \quad (371)$$

$$\geq \mathbb{E}_\mu(X) - 2 + 2\mathbb{E}_\mu\left(\frac{X}{1+X}\right) \quad (372)$$

$$\geq \mathbb{E}_\mu(X) - 2 + 2\mathbb{E}_\mu\left(\frac{X}{1+a}\right) \quad (373)$$

$$\geq -2 = \log q_0(2). \quad (374)$$

3) By our assumption that $f(M)/q_0(M) \geq e$, we have

$$\int_M^\infty f \log f/q_0 \geq \int_M^\infty f. \quad (375)$$

Combining the bounds in the above three parts, we obtain the desired result. \square

APPENDIX F OPTIMIZATION PROBLEM IN (82)

A. Prokhorov's Theorem

Theorem F.1. *Let $\{\mu_n\}$ be a sequence of tight probability measures on \mathbb{R} , i.e., for all $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}$ such that for all $n \geq 1$, $\mu_n(\mathbb{R} \setminus K) \leq \epsilon$. Then, there exists a sub-sequence $\{\mu_{n_k}\}_{k \geq 1}$ and another probability measure μ on \mathbb{R} such that $\{\mu_{n_k}\}_{k \geq 1}$ converges weakly to μ .*

B. Convex Optimization for General Vector Spaces

Theorem F.2. ([27, Theorem 1, Page 217]). Let \mathcal{V} be a vector space, $\Omega \subset \mathcal{V}$ a convex set, \mathcal{U} be a normed vector space, and $\mathcal{P} \subset \mathcal{U}$ be a positive cone, i.e., for all $u_1, u_2 \in \mathcal{U}$ and all $\alpha, \beta \geq 0$, we have $\alpha u_1 + \beta u_2 \in \mathcal{P}$. Suppose the interior of \mathcal{P} is non-empty, and $\phi : \Omega \rightarrow \mathbb{R}$ and $G : \Omega \rightarrow \mathcal{U}$ are convex functions such that there exists $\omega_1 \in \Omega$ for which $G(\omega_1) \prec_{\mathcal{P}} 0$ and $A \triangleq \inf_{\omega \in \Omega: G(\omega) \preceq_{\mathcal{P}} 0} \phi(\omega) > -\infty$. Then, there exists $u_0^* \succeq_{\mathcal{P}^*} 0$ in \mathcal{U}^* such that $A = \inf_{\omega \in \Omega} \phi(\omega) + \langle G(\omega), u_0^* \rangle$. Moreover, if ω_0 is a solution to the first optimization problem, the infimum of the second optimization problem is also achieved by ω_0 and $\langle G(\omega_0), u_0^* \rangle = 0$.

We next recall a result from [16] to find an expression for the KKT conditions of an abstract convex optimization. To this end, we introduce the notation of weak differentiability for a function $f : \Omega \rightarrow \mathbb{R}$ where Ω is convex. We say that $f'_{\omega_0} : \Omega \rightarrow \mathbb{R}$ is the weak derivative of f at ω_0 , if

$$f'_{\omega_0}(\omega) = \lim_{\theta \rightarrow 0^+} \frac{f(\theta\omega + (1-\theta)\omega_0)}{\theta}. \quad (376)$$

Theorem F.3 ([16]). Let \mathcal{V} be a linear space, $\Omega \subset \mathcal{V}$ be convex, and $f : \Omega \rightarrow \mathbb{R}$ be convex and have weak derivative for all $\omega \in \Omega$. $f(\omega^*) = \inf_{\omega \in \Omega} f(\omega)$ if and only if for all $\omega \in \Omega$, we have $f'_{\omega^*}(\omega) \geq 0$.

C. Technical Results

Lemma F.1. $A(\nu)$ defined in (82) satisfies the following properties.

- 1) It is concave and non-decreasing on $[0, \infty)$.
- 2) It is continuous on $[0, \infty)$.
- 3) The one-sided derivatives,

$$A'(\nu^+) \triangleq \lim_{h \rightarrow 0^+} \frac{A(\nu+h) - A(\nu)}{h} \text{ and } A'(\nu^-) \triangleq \lim_{h \rightarrow 0^+} \frac{A(\nu) - A(\nu-h)}{h}, \quad (377)$$

exist for all $\nu > 0$, and for all $0 < \nu_1 < \nu_2$, we have $A'(\nu_1^-) \geq A'(\nu_1^+) \geq A'(\nu_2^-) \geq A'(\nu_2^+)$.

- 4) There exist constants $\nu_0 > 0$ and $C > 0$ such that for all $0 < \nu \leq \nu_0$, we have $A(\nu) \geq C\sqrt{\nu}$.
- 5) We have $\lim_{\nu \rightarrow 0^+} A'(\nu^+) = \lim_{\nu \rightarrow 0^+} A'(\nu^-) = \infty$.

Proof. 1) By definition of $A(\nu)$, it follows that $A(\nu)$ is non-decreasing. To check concavity, we take any $\nu_1, \nu_2 > 0$, $\mu_1, \mu_2 \in \Omega$ with $\mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0) \leq \nu_1$ and $\mathbb{D}(w_{Z|X} \circ \mu_2 \| q_0) \leq \nu_2$, and $\lambda \in [0, 1]$. By convexity of the relative entropy, we have

$$\mathbb{D}(w_{Z|X} \circ (\lambda\mu_1 + (1-\lambda)\mu_2) \| q_0) \leq \lambda\nu_1 + (1-\lambda)\nu_2. \quad (378)$$

Therefore, by concavity of the mutual information,

$$A(\lambda\nu_1 + (1-\lambda)\nu_2) \geq I(\lambda\mu_{\nu_1}^* + (1-\lambda)\mu_{\nu_2}^*, w_{Y|X}) \quad (379)$$

$$\geq \lambda I(\mu_1, w_{Y|X}) + (1-\lambda)I(\mu_2, w_{Y|X}). \quad (380)$$

Hence, by definition of supremum, we have

$$A(\lambda\nu_1 + (1-\lambda)\nu_2) \geq \sup_{\mu_1, \mu_2 \in \Omega: \mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0) \leq \nu_1, \mathbb{D}(w_{Z|X} \circ \mu_2 \| q_0) \leq \nu_2} \lambda I(\mu_1, w_{Y|X}) + (1-\lambda)I(\mu_2, w_{Y|X}) \quad (381)$$

$$= \lambda \sup_{\mu_1 \in \Omega: \mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0) \leq \nu_1} I(\mu_1, w_{Y|X}) + (1-\lambda) \sup_{\mu_2 \in \Omega: \mathbb{D}(w_{Z|X} \circ \mu_2 \| q_0) \leq \nu_2} I(\mu_2, w_{Y|X}) \quad (382)$$

$$= \lambda A(\nu_1) + (1-\lambda)A(\nu_2). \quad (383)$$

- 2) Since $A(\nu)$ is concave on $[0, \infty)$, it is continuous on $(0, \infty)$ [28, Page 153, Problem 4]. To check the continuity at 0, we consider $\nu > 0$ and $\mu \in \Omega$ with $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$. Using (239), we have

$$I(\mu, w_{Y|X}) = - \int_0^\infty (w_{Y|X} \circ \mu)(y) \log(w_{Y|X} \circ \mu)(y) dy - \mathbb{E}_\mu(\log(1 + \theta_m^2 X)) - 1. \quad (384)$$

Furthermore, since $\mathbb{E}_{w_{Y|X} \circ \mu}(Y) = 1 + \theta_m^2 \mathbb{E}_\mu(X)$ by (207) and the support of $w_{Y|X} \circ \mu$ is included in $[0, \infty)$, the differential entropy of $w_{Y|X} \circ \mu$ is upper-bounded by the differential entropy of an exponential distribution with the same mean [29]. Therefore, we have

$$I(\mu, w_{Y|X}) \leq 1 + \log(1 + \theta_m^2 \mathbb{E}_\mu(X)) - \mathbb{E}_\mu(\log(1 + \theta_m^2 X)) - 1 \quad (385)$$

$$\leq \theta_m^2 \mathbb{E}_\mu(X). \quad (386)$$

Furthermore, we have

$$\nu \geq \mathbb{D}(w_{Z|X} \circ \mu \| q_0) \quad (387)$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \quad (388)$$

$$\stackrel{(a)}{=} \int_0^\infty (w_{Z|X} \circ \mu)(z) \log((w_{Z|X} \circ \mu)(z)) dz + \mathbb{E}_{w_{Z|X} \circ \mu}(Z) \quad (389)$$

$$\geq -1 - \log(\mathbb{E}_{w_{Z|X} \circ \mu}(Z)) + \mathbb{E}_{w_{Z|X} \circ \mu}(Z) \quad (390)$$

$$\stackrel{(b)}{=} -\log(1 + \mathbb{E}_\mu(X)) + \mathbb{E}_\mu(X) \quad (391)$$

$$\stackrel{(c)}{\geq} \frac{1}{2}\mathbb{E}_\mu(X)^2 - \frac{1}{3}\mathbb{E}_\mu(X)^3 \quad (392)$$

$$\stackrel{(d)}{\geq} \frac{1}{2}\mathbb{E}_\mu(X)^2 \left(1 - \frac{2}{3}(2\sqrt{\nu} + \nu)\right) \quad (393)$$

where (a) follows since $\log q_0(z) = -z$ and $\mathbb{E}_{w_{Z|X} \circ \mu}(Z) < \infty$ by Lemma C.1 and (208), (b) follows from (208), (c) follows from $\log(1+x) \leq x - x^2/2 + x^3/3$ for $x > -1$, and (d) follows from Lemma C.2. We obtain for $\nu < 1/4$ that $\mathbb{E}_\nu(X) \leq \frac{\sqrt{2\nu}}{1-(2/3)(2\sqrt{\nu}+\nu)} \leq \frac{\sqrt{2\nu}}{1-2\sqrt{\nu}}$, and hence,

$$I(\mu, w_{Y|X}) \leq \frac{\theta_m^2 \sqrt{2\nu}}{1-2\sqrt{\nu}}. \quad (394)$$

Additionally, since $A(\nu)$ is non-decreasing and non-negative, we have

$$|A(\nu) - A(0)| = A(\nu) - A(0) \leq A(\nu) \leq \frac{\theta_m^2 \sqrt{2\nu}}{1-2\nu}, \quad (395)$$

which implies that $A(\nu)$ is continuous at zero.

3) Follows from [28, Page 153, Problem 4] and concavity of $A(\nu)$.

4) For $\nu > 0$ small enough, it is enough to find a probability measure μ satisfying $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ and $I(\mu, w_{Y|X}) \geq C\sqrt{\nu}$. Let μ be a discrete probability measure on \mathcal{X} with two mass points at 0 and \tilde{x} with probabilities $1 - \alpha$ and α , respectively, such that $\tilde{x} < \min(1, 1/\theta_m^2)$. Then, by Lemma C.6,

$$\mathbb{D}(w_{Z|X} \circ \mu \| q_0) = \frac{\alpha^2 \tilde{x}^2}{2(1 - \tilde{x}^2)} + o(\alpha^2). \quad (396)$$

Similarly, we can obtain $\mathbb{D}(w_{Y|X} \circ \mu \| p_0) \leq \alpha^2 \theta_m^2 \tilde{x}^2 / (2(1 - \theta_m^2 \tilde{x}^2)) + o(\alpha^2)$. Therefore, we can lower-bound the mutual information by

$$I(\mu, w_{Y|X}) = \alpha \mathbb{D}(p_{\tilde{x}} \| p_0) - \mathbb{D}(w_{Y|X} \circ \mu \| p_0) \quad (397)$$

$$\geq \alpha \mathbb{D}(p_{\tilde{x}} \| p_0) - \frac{\alpha^2 \theta_m^2 \tilde{x}^2}{2(1 - \theta_m^2 \tilde{x}^2)} - o(\alpha^2) \quad (398)$$

$$= \alpha (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})) - \frac{\alpha^2 \theta_m^2 \tilde{x}^2}{2(1 - \theta_m^2 \tilde{x}^2)} - o(\alpha^2). \quad (399)$$

Hence, by choosing $\alpha = \tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2) \mathbb{D}(w_{Z|X} \circ \mu \| q_0)} = \tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2) \nu (1 - o(1))}$, we have $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$ and

$$I(\mu, w_{Y|X}) \geq \sqrt{\nu} (1 - o(1)) \left(\tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2)} (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})) - \sqrt{\nu} \frac{1 - \tilde{x}^2}{1 - \theta_m^2 \tilde{x}^2} \right). \quad (400)$$

Choosing $\nu_0 > 0$ such that

$$\tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2)} (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})) \geq 2\sqrt{\nu_0} \frac{1 - \tilde{x}^2}{1 - \theta_m^2 \tilde{x}^2}, \quad (401)$$

the claim of the lemma holds for

$$C = \frac{1}{2} \tilde{x}^{-1} \sqrt{2(1 - \tilde{x}^2)} (\theta_m^2 \tilde{x} - \log(1 + \theta_m^2 \tilde{x})). \quad (402)$$

5) Since $A'(\nu^+) \leq A'(\nu^-)$, we only need to compute $\lim_{\nu \rightarrow 0^+} A'(\nu^+)$. Since $A(\nu)$ is concave $A'(\nu^+)$ is decreasing, and therefore, it is enough to show that for any $L > 0$ there exists some $\nu > 0$ with $A'(\nu^+) \geq L$. To this end, we fix some $\tilde{\nu} > 0$ and define $B(\nu) \triangleq A(\nu) - \frac{A(\tilde{\nu})}{\tilde{\nu}} \nu$. $B(\nu)$ is continuous on $[0, \tilde{\nu}]$ and therefore it achieves its maximum and minimum

on $[0, \tilde{\nu}]$. Hence, either we have $B(\nu) = 0$ for all $\nu \in [0, \tilde{\nu}]$ or there exists a $\nu \in (0, \tilde{\nu})$ such that $B(\nu)$ achieves its maximum or minimum at ν . Then, we should have $B'(\nu^-) = A'(\nu^-) - A(\tilde{\nu})/\tilde{\nu} \geq 0$ or $B'(\nu^+) = A'(\nu^+) - A(\tilde{\nu})/\tilde{\nu} \geq 0$. In both cases, we have $A'(\frac{\nu^+}{2}) \geq \frac{A(\tilde{\nu})}{\tilde{\nu}}$. However, by Lemma F.1, $A'(\frac{\nu^+}{2}) \geq C/\sqrt{\tilde{\nu}}$, if $\tilde{\nu} \leq \nu_0$. Since $\tilde{\nu}$ is arbitrary, we can choose it such that $C/\sqrt{\tilde{\nu}} > L$. \square

Proof of Lemma IV.3. We only prove the existence of a solution and the uniqueness follows from strict concavity of the mutual information [18]. Consider a sequence $\{\mu_n\}_{n \geq 1}$ in Ω such that $\mathbb{D}(w_{Z|X} \circ \mu_n \| q_0) \leq \nu$ and $\lim_{n \rightarrow \infty} I(\mu_n, w_{Y|X}) = A(\nu)$. To use F.1, we first check that this sequence is tight. For any $\epsilon > 0$, we have

$$\mathbb{P}_{\mu_n}(X \notin [0, (2\sqrt{\nu} + \nu)/\epsilon]) \stackrel{(a)}{\leq} \frac{\mathbb{E}_{\mu_n}(X)\epsilon}{2\sqrt{\nu} + \nu} \quad (403)$$

$$\stackrel{(b)}{\leq} \epsilon, \quad (404)$$

where (a) follows from applying Markov's inequality to the almost surely non-negative random variable X , and (b) follows from Lemma C.2. Since $[0, (2\sqrt{\nu} + \nu)/\epsilon]$ is compact, the sequence $\{\mu_n\}_{n \geq 1}$ is tight. Therefore, we are permitted to use Theorem F.1 that shows the existence of a subsequence $\{\mu_{n_k}\}_{k \geq 1}$ and probability measure μ on \mathbb{R} such that $\{\mu_{n_k}\}_{k \geq 1}$ converges weakly to μ . We claim that $\mu_{n_k}^*$ is indeed μ and prove it in three steps.

Step 1: Theorem F.1 only guarantees the existence of a probability measure on \mathbb{R} which can possibly have positive measure on negative numbers. In this step, we show that this is not the case. By the Portmanteau theorem, the weak convergence of $\{\mu_{n_k}\}_{k \geq 1}$ to μ implies that $\liminf_{k \rightarrow \infty} \mu_{n_k}(U) \geq \mu(U)$ for any open set $U \subset \mathbb{R}$. Taking $U =] - \infty, 0[$, we obtain that

$$0 = \liminf_{k \rightarrow \infty} \mu_{n_k}(] - \infty, 0[) \geq \mu(] - \infty, 0[) \geq 0, \quad (405)$$

which means that $\mu(] - \infty, 0[) = 0$.

Step 2: In this step we prove that μ satisfies the optimization constraint, i.e., $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) \leq \nu$. Let us define $f_k(z) \triangleq (w_{Z|X} \circ \mu_{n_k})(z)$ and $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$. Since for any $z \in \mathcal{Z}$, $q_x(z) = e^{-z/(1+x)}/(1+x)$ is a continuous and bounded function in x , by weak convergence definition, we have

$$f_k(z) = \mathbb{E}_{\mu_{n_k}}(q_X(z)) \rightarrow \mathbb{E}_{\mu}(q_X(z)) = f(z). \quad (406)$$

In the next lemma, we show that $|f_k(z) \log f_k(z)|$ is uniformly upper-bounded by an integrable function.

Lemma F.2. *There exists some \tilde{z} such that for all k ,*

$$|f_k(z) \log f_k(z)| \leq g(z) \triangleq \begin{cases} e^{-1} & z \in [0, \tilde{z}], \\ \frac{2\sqrt{\nu} + \nu}{e(z^{\frac{3}{2}} - z^{\frac{1}{2}})} + z^{-\frac{1}{2}} e^{-\sqrt{z}} & z \in [\tilde{z}, \infty[, \end{cases} \quad (407)$$

and $\int_0^\infty |g(z)| dz < \infty$.

Proof. Note first that for all $x \in [0, 1]$, we have $|x \log x| \leq e^{-1}$, and for all $x \in [0, e^{-1}]$, we have $|x \log x| \leq |x|$. Thus, it is enough to show that there exist \tilde{z} such that for all $k \geq 1$ and $z \geq \tilde{z}$,

$$f_k(z) \leq \frac{2\sqrt{\nu} + \nu}{e(z^{\frac{3}{2}} - z^{\frac{1}{2}})} + z^{-\frac{1}{2}} e^{-\sqrt{z}}. \quad (408)$$

By law of total probability, for all $\lambda > 0$, we have

$$f_k(z) \triangleq \mathbb{E}_{\mu_{n_k}}(q_X(z)) \quad (409)$$

$$= \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geq \lambda) \mathbb{P}_{\mu_{n_k}}(X \geq \lambda) + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda) \mathbb{P}_{\mu_{n_k}}(X < \lambda) \quad (410)$$

$$\stackrel{(a)}{\leq} \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geq \lambda) \frac{\mathbb{E}_{\mu_{n_k}}(X)}{\lambda} + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda) \quad (411)$$

$$\stackrel{(b)}{\leq} \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geq \lambda) \frac{2\sqrt{\nu} + \nu}{\lambda} + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda), \quad (412)$$

where (a) follows from Markov's inequality, and (b) follows from Lemma C.2. We also have for all $z \geq 1$, $q_x(z) \leq (ze)^{-1}$, and for all $0 \leq x \leq \lambda \leq z - 1$, $q_x(z) \leq e^{-\frac{z}{1+x}}/(1+x)$. Substituting these upper-bounds in (412) for $\lambda = z^{\frac{1}{2}} - 1$, which is less than $z - 1$ for $z \geq 1$, we obtain

$$f_k(z) \leq \frac{1}{ze} \frac{2\sqrt{\nu} + \nu}{z^{\frac{1}{2}} - 1} + \frac{1}{z^{\frac{1}{2}}} e^{-z^{\frac{1}{2}}} \quad (413)$$

$$= \frac{2\sqrt{\nu} + \nu}{e(z^{\frac{3}{2}} - z^{\frac{1}{2}})} + z^{-\frac{1}{2}} e^{-\sqrt{z}}. \quad (414)$$

□

We are now eligible to use dominated convergence theorem and exchange limit and integral to obtain

$$\lim_{k \rightarrow \infty} \int_0^\infty f_k(z) \log f_k(z) dz = \int_0^\infty \lim_{k \rightarrow \infty} f_k(z) \log f_k(z) dz \quad (415)$$

$$= \int_0^\infty f(z) \log f(z) dz. \quad (416)$$

Since $f_k(z)z \geq 0$ for all $z \in \mathcal{Z}$ and $k \geq 1$, Fatou's lemma yields that

$$\int_0^\infty f(z)z dz = \int_0^\infty \liminf_{k \rightarrow \infty} f_k(z)z dz \quad (417)$$

$$\leq \liminf_{k \rightarrow \infty} \int_0^\infty f_k(z)z dz. \quad (418)$$

Combing (416) and (418), we have

$$\mathbb{D}(w_{Z|X} \circ \mu \| q_0) = \int_0^\infty f(z) (z + \log f(z)) dz \quad (419)$$

$$\leq \liminf_{k \rightarrow \infty} \int_0^\infty f_k(z) (z + \log f_k(z)) dz = \liminf_{k \rightarrow \infty} \mathbb{D}(w_{Z|X} \circ \mu_{n_k} \| q_0) \leq \nu. \quad (420)$$

Step 3: It remains to show that $I(\mu, w_{Y|X}) \geq A(\nu)$. We again define $h_k(z) \triangleq (w_{Y|X} \circ \mu_{n_k})(z)$ and $h(z) \triangleq (w_{Y|X} \circ \mu)(z)$. With the same argument of the previous step, we can prove that

$$\lim_{k \rightarrow \infty} \int_0^\infty h_k(z) \log h_k(z) dz = \int_0^\infty h(z) \log h(z) dz. \quad (421)$$

Furthermore, by [30, Page 86], we have

$$\liminf_{k \rightarrow \infty} \mathbb{E}_{\mu_{n_k}} (1 + \log(1 + \theta_m^2 X)) \geq \mathbb{E}_\mu (1 + \log(1 + \theta_m^2 X)). \quad (422)$$

Hence, (239) implies that $I(\mu, w_{Y|X}) \geq A(\nu)$. □

Proof of Theorem IV.1. We prove all four statements in order. The proof heavily relies on results from convex optimization for general vector spaces and properties of the optimization problem in (83), which we have gathered in Appendix F for the reader's convenience.

- 1) In Theorem F.2, taking Ω as the set of all probability measures μ on \mathcal{X} with $\mathbb{D}(w_{Z|X} \circ \mu \| q_0) < \infty$, $\mathcal{U} = \mathbb{R}$, $\mathcal{P} = \mathbb{R}^+$, $\phi(\mu) = -I(\mu, w_{Y|X})$, $G(\mu) = \mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu$, we note that

$$-\infty < -A(\nu) = - \sup_{\mu \in \Omega: G(\mu) \leq 0} -\phi(\mu) \quad (423)$$

$$= \inf_{\mu \in \Omega: G(\mu) \leq 0} \phi(\mu). \quad (424)$$

By convexity of the relative entropy and concavity of mutual information in the input distribution, ϕ and G are convex functions, with μ_1 the deterministic probability measure with all mass point at zero, we also have $G(\mu_1) = -\nu < 0$. Therefore, we can apply Theorem F.2 to show the existence of $\gamma(\nu) \geq 0$ such that

$$\inf_{\mu \in \Omega: G(\mu) \leq 0} \phi(\mu) = \inf_{\mu \in \Omega} [\phi(\mu) + \gamma(\nu)G(\mu)] \quad (425)$$

$$= - \sup_{\mu \in \Omega} [I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)], \quad (426)$$

which results in the unconstrained reformulation of $A(\nu)$ as $\sup_{\mu \in \Omega} [I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)]$. Theorem F.2 also implies that μ_ν^* is a solution to this new optimization problem, and since $I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)$ is strictly concave [18, Appendix I.B], the solution is unique.

- 2) With the help of Lemma F.3 in Appendix C to show the existence of weak derivatives (defined in (376)), we use Theorem F.3 with $f(\mu) = I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)$ to obtain that $\mu_1 = \mu_\nu^*$ if and only if for any $\mu \in \Omega$,

$$0 \geq f'_{\mu_1}(\mu) \quad (427)$$

$$\begin{aligned} &= \mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu_1)(Y)} \right) - I(\mu_1, w_{Y|X}) \\ &\quad - \gamma(\nu) \left(\mathbb{E}_{w_{Z|X} \circ \mu} \left(\log \frac{(w_{Z|X} \circ \mu_1)(Z)}{q_0(Z)} \right) - \mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0) \right) \end{aligned} \quad (428)$$

$$= \mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu_1)(Y)} \right) - \gamma(\nu) \left(\mathbb{E}_{w_{Z|X} \circ \mu} \left(\log \frac{(w_{Z|X} \circ \mu_1)(Z)}{q_0(Z)} \right) - \nu \right) - f(\mu_1) \quad (429)$$

$$= \mathbb{E}_\mu(w(X, \mu_1, \nu)) - f(\mu_1). \quad (430)$$

This implies that $\mu_1 = \mu_\nu^*$ if and only if for all $\mu \in \Omega$, we have $f(\mu_1) \geq \mathbb{E}_\mu(w(X, \mu_1, \nu))$. Since $A(\nu) = \sup_{\mu \in \Omega} f(\mu) \geq f(\mu_1)$, if $\mu_1 = \mu_\nu^*$, then for all $\mu \in \Omega$, we have $A(\nu) \geq f(\mu_1) \geq \mathbb{E}_\mu(w(X, \mu_1, \nu))$.

- 3) Assume (87) is true, we take the expectation and obtain (85). We now show the opposite direction and prove that if (85) holds, we have (87) and (88). Applying (85) with μ a deterministic probability measure with all mass point at x , we obtain

$$A(\nu) \geq \mathbb{E}_\mu(w(X, \mu_1, \nu)) = w(x, \mu_1, \nu). \quad (431)$$

Furthermore, for any $x \in \text{support}(\mu_1)$, we prove that $w(x, \mu_1, \nu) = A(\nu)$ by contradiction. If $A(\nu) - w(x, \mu_1, \nu) \triangleq \delta > 0$, by continuity of $A(\nu) - w(x, \mu_1, \nu)$ in x , there exists a neighborhood \mathcal{N} of x such that for all $x' \in \mathcal{N}$, we have $A(\nu) - w(x', \mu_1, \nu) \geq \delta/2$. Also, since $x \in \text{support}(\mu_1)$, we know that $\mathbb{P}_{\mu_1}(X \in \mathcal{N}) = \epsilon > 0$. Therefore, we obtain

$$A(\nu) = \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)) = \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)\mathbb{1}\{X \in \mathcal{N}\}) + \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)\mathbb{1}\{X \notin \mathcal{N}\}) \quad (432)$$

$$\leq (1 - \epsilon)A(\nu) + \epsilon \left(A(\nu) - \frac{\delta}{2} \right) \quad (433)$$

$$= A(\nu) - \frac{\delta\epsilon}{2} < A(\nu), \quad (434)$$

which is a contradiction.

- 4) To prove that $\lim_{\nu \rightarrow 0^+} \gamma(\nu) = \infty$, we prove that $\gamma(\nu) \geq A'(\nu^+)$, and the result will follow from $\lim_{\nu \rightarrow 0^+} A'(\nu^+) = \infty$ as shown in Lemma F.1. Consider any $\nu_1, \nu_2 > 0$, and similar to the sensitivity analysis in [19, Section 5.6], note that

$$A(\nu_1) = I(\mu_{\nu_1}^*, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}(w_{Z|X} \circ \mu_{\nu_1}^* \| q_0) - \nu_1) \quad (435)$$

$$\stackrel{(a)}{\geq} I(\mu_{\nu_2}^*, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}(w_{Z|X} \circ \mu_{\nu_2}^* \| q_0) - \nu_1) \quad (436)$$

$$= I(\mu_{\nu_2}^*, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}(w_{Z|X} \circ \mu_{\nu_2}^* \| q_0) - \nu_2) + \gamma(\nu_1)(\nu_1 - \nu_2) \quad (437)$$

$$\stackrel{(b)}{\geq} I(\mu_{\nu_2}^*, w_{Y|X}) + \gamma(\nu_1)(\nu_1 - \nu_2), \quad (438)$$

where (a) follows since $\mu_{\nu_1}^*$ is the maximizer of $\sup_{\mu} I(\mu, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu_1)$, and (b) follows since $\gamma(\nu_1) \geq 0$. Thus, for any $\nu > 0$ and $\nu > h > 0$, we have

$$\frac{A(\nu) - A(\nu - h)}{h} \geq \gamma(\nu) \quad \text{and} \quad \frac{A(\nu + h) - A(\nu)}{h} \leq \gamma(\nu). \quad (439)$$

Taking the limit $h \rightarrow 0^+$, we obtain $A'(\nu^+) \leq \gamma(\nu) \leq A'(\nu^-)$.

To prove that $\lim_{\nu \rightarrow 0^+} \gamma(\nu)\nu = 0$, note that for all $\nu > 0$,

$$\gamma(\nu)\nu \leq A'(\nu^-)\nu \quad (440)$$

$$\stackrel{(a)}{\leq} \frac{A(\nu)}{\nu}\nu = A(\nu), \quad (441)$$

where (a) follows from concavity of A . In the proof of Lemma F.1, we show that $\lim_{\nu \rightarrow 0^+} A(\nu) = 0$, which yields the result. \square

Lemma F.3. $f(\mu) \triangleq I(\mu, w_{Y|X}) - \gamma(\nu)(\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)$ is weakly differentiable, and

$$f'_{\mu_1}(\mu) = \mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu_1)(Y)} \right) - I(\mu_1, w_{Y|X}) - \gamma(\nu) \left(\mathbb{E}_{w_{Z|X} \circ \mu} \left(\log \frac{(w_{Z|X} \circ \mu_1)(Z)}{q_0(Z)} \right) - \mathbb{D}(w_{Z|X} \circ \mu_1 \| q_0) \right). \quad (442)$$

Proof. In [18, Equation (63)], the weak derivative of $I(\mu, w_{Y|X})$ at μ_1 is proved to be

$$\mathbb{E}_{w_{Y|X} \times \mu} \left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu_1)(Y)} \right) - I(\mu_1, w_{Y|X}). \quad (443)$$

Thus, we only check the weak differentiability of $G(\mu) \triangleq \mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu$. Let $\mu_1, \mu \in \Omega$, and define

$$\mu_\theta \triangleq (1 - \theta)\mu_1 + \theta\mu \quad (444)$$

$$f_1(z) \triangleq (w_{Z|X} \circ \mu_1)(z) \quad (445)$$

$$f(z) \triangleq (w_{Z|X} \circ \mu)(z) \quad (446)$$

$$f_\theta(z) \triangleq (w_{Z|X} \circ \mu_\theta)(z). \quad (447)$$

Then, we have

$$G(\mu_\theta) - G(\mu_1) \quad (448)$$

$$= \mathbb{D}(f_\theta \| q_0) - \mathbb{D}(f_1 \| q_0) \quad (449)$$

$$= \int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz \quad (450)$$

$$\stackrel{(a)}{=} \int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{q_0(z)} dz - \int_0^\infty f_\theta(z) \log \frac{f_1(z)}{q_0(z)} dz + \int_0^\infty f_\theta(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz \quad (451)$$

$$= \int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} dz + \int_0^\infty f_\theta(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz \quad (452)$$

$$\stackrel{(b)}{=} \int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} dz + \theta \left(\int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz \right), \quad (453)$$

where (a) holds since by Lemma C.4, $\int_0^\infty f_\theta(z) \log \frac{f_1(z)}{q_0(z)} dz < \infty$, and (b) follows from $f_\theta = (1 - \theta)f_1 + \theta f$. The second term in (453) is differentiable with respect to θ , and the derivative is

$$\int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz. \quad (454)$$

To take derivative from the first term in (453), we use Theorem A.1. Note that by Lemma C.4, $\int_0^\infty f_\theta(z) \left| \log \frac{f_\theta(z)}{f_1(z)} \right| dz < \infty$, and also, for all z and θ ,

$$\frac{\partial}{\partial \theta} \left(f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} \right) = -(f_1(z) - f(z)) \left(1 + \log \frac{f_\theta(z)}{f_1(z)} \right). \quad (455)$$

Additionally, for all $\theta \in [0, 1]$, if we apply (206), we obtain

$$|f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)}| \leq |f_1(z) + f(z)| (|\log f_1(z)| + \log(1 + \mathbb{E}_{\mu_\theta}(X)) + z) \quad (456)$$

$$\leq |f_1(z) + f(z)| (|\log f_1(z)| + \log(1 + \mathbb{E}_{\mu_1}(X) + \mathbb{E}_\mu(X)) + z), \quad (457)$$

which is an integrable function with respect to Lebesgue measure on \mathcal{Z} by Lemma C.4 and does not depend on θ . Hence, all condition in Theorem A.1 hold, and we have

$$\frac{\partial}{\partial \theta} \left(\int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} dz \right) = \int_0^\infty \frac{\partial}{\partial \theta} \left(f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} \right) dz \quad (458)$$

$$= \int_0^\infty -(f_1(z) - f(z)) \left(1 + \log \frac{f_\theta(z)}{f_1(z)} \right) dz \quad (459)$$

$$= \int_0^\infty -(f_1(z) - f(z)) \log \frac{f_\theta(z)}{f_1(z)} dz, \quad (460)$$

which vanishes at $\theta = 0$. Therefore, G is weakly differentiable at μ_1 and

$$G'_{\mu_1}(\mu) = \int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz. \quad (461)$$

Since the mutual information and the divergence are weakly differentiable, so is $I(\mu, w_{Y|X}) - \gamma(\nu) (\mathbb{D}(w_{Z|X} \circ \mu \| q_0) - \nu)$. \square

REFERENCES

- [1] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, September 2013.
- [2] B. A. Bash, A. H. Gheorghie, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, no. 1, p. 8626, Dec 2015.
- [3] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013, pp. 2945–2949.

- [4] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.
- [5] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.
- [6] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, p. 21902212, Apr 2019.
- [7] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. of IEEE Information Theory Workshop*, Cambridge, UK, September 2016, pp. 364–368.
- [8] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. of IEEE International Symposium on Information Theory*, Barcelona, Spain, July 2016, pp. 2064–2068.
- [9] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, "Gaussian signalling for covert communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, p. 35423553, Jul 2019.
- [10] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Codes for covert communication over additive white gaussian noise channels," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 977–981.
- [11] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, 2017.
- [12] —, "Covert communications on continuous-time channels in the presence of jamming," in *Proc. of 51st Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2017, pp. 1697–1701.
- [13] B. A. Bash, D. Goeckel, and D. Towsley, "Covert communication gains from adversary's ignorance of transmission time," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8394–8405, Dec. 2016.
- [14] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, "Delay-intolerant covert communications with either fixed or random transmit power," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 129–140, Jan. 2019.
- [15] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4766–4779, Jul. 2018.
- [16] J. G. Smith, "On the information capacity of peak and average power constrained gaussian channels," Ph.D. dissertation, University of California, Berkeley, 1969.
- [17] —, "The information capacity of amplitude and variance-constrained scalar gaussian channels," *Information and Control*, vol. 18, pp. 203–219, 1971.
- [18] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.
- [19] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [20] L. Wang, G. Wornell, and L. Zheng, "Limits of low-probability-of-detection communication over a discrete memoryless channel," in *Proceedings of 2015 IEEE International Symposium on Information Theory*, 2015, pp. 2525–2529.
- [21] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [22] M. Hayashi, "Quantum wiretap channel with non-uniform random number and its exponent and equivocation rate of leaked information," *IEEE Transactions on Information Theory*, vol. 61, no. 10, pp. 5595–5622, 2015.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.
- [24] J. Hou, "Coding for relay networks and effective secrecy for wire-tap channels," Ph.D. dissertation, 2014.
- [25] M. J. Ablowitz and A. S. Fokas, *Complex variables: introduction and applications*. Cambridge University Press, 2003.
- [26] R. B. Ash, *Information Theory*, ser. Interscience Tracts in Pure and Applied Mathematics. John Wiley & Sons, 1965.
- [27] D. G. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1997.
- [28] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2009.
- [29] S. Y. Park and A. K. Bera, "Maximum entropy autoregressive conditional heteroskedasticity model," *Journal of Econometrics*, vol. 150, no. 2, pp. 219–230, 2009.
- [30] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2010.